

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 14, 2001, 07:52:13 ; Search time 2960.24 Seconds
(without alignments)
4561.571 Million cell updates/sec

Title: US-09-105-117I-1_COPY_1421_2293
Perfect score: 873
Sequence: 1 atgaacccattcaactgga.....tcgagggattgcggccttag 873

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues 2688314

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl:*

- 1: gb_ba1:*
- 2: gb_ba2:*
- 3: gb_ba3:*
- 4: gb_in1:*
- 5: gb_in2:*
- 6: gb_in3:*
- 7: gb_cm:*
- 8: gb_ov:*
- 9: gb_pat1:*
- 10: gb_pat2:*
- 11: gb_ph:*
- 12: gb_pl1:*
- 13: gb_pl2:*
- 14: gb_pl3:*
- 15: gb_pl4:*
- 16: em_ba1:*
- 17: em_ba2:*
- 18: em_fun:*
- 19: em_htgo_hum:*
- 20: em_htgo_inv:*
- 21: em_htgo_rod:*
- 22: em_htg_hum1:*
- 23: em_htg_hum2:*
- 24: em_htg_hum3:*
- 25: em_htg_hum4:*
- 26: em_htg_hum5:*
- 27: em_htg_hum6:*
- 28: em_htg_hum7:*
- 29: em_htg_hum8:*
- 30: em_htg_inv1:*
- 31: em_htg_inv2:*
- 32: em_htg_other:*
- 33: em_htg_rod:*
- 34: em_hum1:*
- 35: em_hum2:*
- 36: em_hum3:*
- 37: em_hum4:*
- 38: em_hum5:*
- 39: em_hum6:*
- 40: em_hum7:*
- 41: em_in:*
- 42: em_cm:*
- 43: em_or:*

- 44: em_ov:*
- 45: em_pat:*
- 46: em_ph:*
- 47: em_pl:*
- 48: em_ro:*
- 49: em_sts:*
- 50: em_sy:*
- 51: em_un:*
- 52: em_v1:*
- 53: gb_sts1:*
- 54: gb_sts2:*
- 55: gb_sts3:*
- 56: gb_sy:*
- 57: gb_un:*
- 58: gb_v1:*
- 59: gb_v12:*
- 60: gb_htg1:*
- 61: gb_htg2:*
- 62: gb_htg3:*
- 63: gb_htg4:*
- 64: gb_htg5:*
- 65: gb_htg6:*
- 66: gb_htg7:*
- 67: gb_htg8:*
- 68: gb_htg9:*
- 69: gb_htg10:*
- 70: gb_htg11:*
- 71: gb_htg12:*
- 72: gb_htg13:*
- 73: gb_htg14:*
- 74: gb_htg15:*
- 75: gb_htg16:*
- 76: gb_htg17:*
- 77: gb_htg18:*
- 78: gb_htg19:*
- 79: gb_htg20:*
- 80: gb_htg21:*
- 81: gb_htg22:*
- 82: gb_htg23:*
- 83: gb_htg24:*
- 84: gb_htg25:*
- 85: gb_pr1:*
- 86: gb_pr2:*
- 87: gb_pr3:*
- 88: gb_pr4:*
- 89: gb_pr5:*
- 90: gb_pr6:*
- 91: gb_pr7:*
- 92: gb_pr8:*
- 93: gb_pr9:*
- 94: gb_rol:*
- 95: gb_rol2:*
- 96: gb_in4:*
- 97: gb_pr10:*
- 98: em_ba3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	873	100.0	993	9	AX063767
c	2	873	100.0	2374	2 CGLYSEG
c	3	873	100.0	2374	9 A93933
c	4	140.6	16.1	33285	3 SC5F8
c	5	122.6	14.0	14844	3 MBU34849
c	6	122.6	14.0	15239	10 I86263
c	7	121	13.9	38500	3 MTCY39
c	8	108.6	12.4	12070	1 AB011413

- AX063767 Sequence
- X96471 C.glutamici
- A93933 Sequence 2
- AL37613 Streptomy
- U34849 Mycobacteri
- I86263 Sequence 17
- Z74035 Mycobacteri
- AB011413 Streptomy

9	82.6	9.5	1804	2	ECRPIAA	X73026 E.coli gene
10	82.6	9.5	1815	2	ECOCICA	M62865 E.coli chro
11	82.6	9.5	6055	2	ECSEPAICI	X56836 E.coli shra
12	82.6	9.5	10362	1	AE000375	AE000375 Escherich
13	82.6	9.5	10367	1	AE005521	AE005521 Escherich
14	82.6	9.5	110000	66	AC020850_0	AC020850 Mus muscu
15	82.6	9.5	141744	2	ECU28377	U28377 Escherichia
16	82.6	9.5	266658	2	AP002563	AP002563 Escherich
17	73	8.4	4033	2	ASU65741	U65741 Aeromonas s
18	72.6	8.3	10542	1	AE004852	AE004852 Pseudomon
19	71.8	8.2	6362	1	AF037440	AF037440 Edwardsie
20	49	5.6	11168	1	AE004560	AE004560 Pseudomon
21	45.4	5.2	45613	10	AX067447	AX067447 Sequence
22	45.2	5.2	10572	1	AE004780	AE004780 Pseudomon
23	43.6	5.0	10698	1	AE004134	AE004134 Vibrio ch
24	43.4	5.0	110000	66	AC020884_0	AC020884 Mus muscu
25	43	4.9	14789	1	AE004916	AE004916 Pseudomon
26	42	4.8	4532	3	SYOSRRA	L41665 Synecococc
27	40.6	4.7	10010	1	AE004639	AE004639 Pseudomon
28	38.8	4.4	1033	1	AB053349	AB053349 Synecococ
29	38.8	4.4	4217	2	AY017305	AY017305 Synecococ
30	38.6	4.4	1070	3	PSECATRA	M33817 P.putida ca
31	38.6	4.4	13246	3	MTV003	AL008883 Mycobacte
32	37.8	4.3	11015	1	AE004736	AE004736 Pseudomon
33	37.6	4.3	13193	1	AE005456	AE005456 Escherich
34	37.6	4.3	26857	2	AP002560	AP002560 Escherich
35	37.2	4.3	10614	1	AE004351	AE004351 Vibrio ch
36	37	4.2	1592	3	S80112	S80112 Ralstonia e
37	37	4.2	1790	3	PJPTFDRII	M98445 Ralstonia e
38	37	4.2	2058	2	AFATFDA	M16730 A.eutrophus
39	37	4.2	2058	9	AI2337	AI2337 tfda gene c
40	37	4.2	2058	10	I36756	I36756 Sequence 1
41	37	4.2	3432	2	ASP271623	AJ271623 Anabaena
42	36.8	4.2	2936	3	KPNACP	L01114 Klebsiella
43	36.8	4.2	11489	1	AE002455	AE002455 Neisseria
44	36.8	4.2	331801	3	NM4422491	AL182735 Neisseria
45	36.8	4.2	349116	2	AP003003	AP003003 Mesorhizo

ALIGNMENTS

RESULT 1	AX063767	Sequence 49 from Patent WO0100843.	993 bp	DNA	PAT	24-JAN-2001
LOCUS	AX063767	AX063767.1	GI:12541479			
DEFINITION	AX063767					
ACCESSION	AX063767					
VERSION	AX063767.1					
KEYWORDS						
SOURCE						
ORGANISM						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
FEATURES						
source						
CDS						

BASE COUNT	233 a	256 c	268 g	236 t	
ORIGIN					
Query Match	100.0%	Score 873;	DB 9;	Length 993;	
Best Local Similarity	100.0%	Pred. No. 1.2e-235;			
Matches 873;	Conservative	0;	Mismatches	0;	Gaps
0;					
QY	1	atgaacccattcaatgacacatttgcctcattgaatgaagcagcttcgaagc	60		
Db	101	ATGAACCCATTCACATGGACACATTGCTCTCAATGATGAAGCAGCTTCGAAGC	160		
QY	61	gcctcttagccttccattccctcctgcgctgagtcagcgttaaacgtctcag	120		
Db	161	GCCTCTTAGCCCTTTCCATTTCCCTCGCGGTGAGTCAGCGGTAAAGCTCTCGAG	220		
QY	121	catcagtggtgcagtggtgtatcgcgaccccaacggcccaaacgcaagcggt	180		
Db	221	CATCAGTGGTGTGAGTGTGTATCGCGACCCCAACCGCCCAAGCAACCGAAGCGGT	280		
QY	181	gaagtcctgtgcaagcagcgcggaataatggtgtgctgcaagcaataaagc	240		
Db	281	GAAGTCCTTGTGAAGCAGCGCGGAAATGCTGTGTCGAAGCAAACTAAAGCGCA	340		
QY	241	ctatctggagccttctgaataatcccttaaccatcgcacatcaacagcagcttc	300		
Db	341	CTATCTGGAGCGCTTGTGTAATCCCGTTAAACCATGCCATCGCAGATTCGCTATCC	400		
QY	301	acatggttctcctccgtgttcaacaggttagcttcttgggtggagcaacgcctg	360		
Db	401	ACATGTTTCTCCCGTGTTCACAGAGGTAGTCTTGTGGGTGGAGCAACGCTCAGCTG	460		
QY	361	cgcttggaagatgaagcagcacattatccttgcggtgagtgatgttttagagcg	420		
Db	461	CGCTTGAAGATGAAGCGCACATATCTCTGCTCGCGGTGGAGATGTTTGAAGCG	520		
QY	421	gtaacccgtgaagctaataccgtggtggtggtggtggtggtggtggtggtggtg	480		
Db	521	GTAAACCGTGAAGCTAATCCCGTGGCGGGATGTGAAGTAGTAGTAACCTTGAACCATGGCG	580		
QY	481	cacttgccattgcaacccctcattgctggtggtggtggtggtggtggtggtggtg	540		
Db	581	CACCTGGCCATTGCAACCCCTCATTCGCGGATGCTTACATGTTGATGGGAACTAGAT	640		
QY	541	tgggtgcgagtcgcttctttagcttgcgtcccaagagatgcttcaaacctgacctg	600		
Db	641	TGGGCTGCGATGCCCTTCTTACGCTTCCGTTCCCAAGATGTGCTTCAAGACCGTGACCTG	700		
QY	601	gacggcggtgctgctgctggtggtggtggtggtggtggtggtggtggtggtggtg	660		
Db	701	GACGGCGGTGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	760		
QY	661	ggtttggtgaggaattgcgcgagccttgcgtggtggtggtggtggtggtggtggtg	720		
Db	761	GGTTTGGTGAAGCAATTCGCCGAGGCTTGGTGGGACTTCTTCCGAAACCAAGCT	820		
QY	721	gtcccatgctaaagcaggaagtgatcctcctcgatgagatgataccattgacacag	780		
Db	821	GCTCCCATGCTAAAGCAGGAGAGTATCTCTCTCGATGAGATGCCATTTGACACACG	880		
QY	781	atgtattgcaacgagtggtggtggtggtggtggtggtggtggtggtggtggtggtg	840		
Db	881	ATGATTGGCAACGATGGCGCTTGAATCTAGATCTCTAGTACTACAGACGCGCTC	940		
QY	841	gttgatgcagcaatcagaggtggtggtggtggtggtggtggtggtggtggtggtggtg	900		
Db	941	GTTGATGCAGCAATCAGAGGATTGCGGCTTAG	973		

RESULT 2

```

CGLYSEG/c      CGLYSEG      2374 bp      DNA      BCT      18-MAR-2001
LOCUS          C.glutamicum lysE and lysG genes.
DEFINITION
ACCESSION      X96471
VERSION        GI:1729753
KEYWORDS       lysE gene; lysG gene; Lysine export regulator protein; Lysine
SOURCE
ORGANISM       Corynebacterium glutamicum.
               Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
               Actinomycetales; Corynebacterineae; Corynebacteriaceae;
               Corynebacterium.
REFERENCE      1 (bases 1 to 2374)
AUTHORS        Vrijic,M., Sahm,H. and Eggeling,L.
TITLE          A new type of transporter with a new type of cellular function:
               L-lysine export from Corynebacterium glutamicum
JOURNAL        Mol. Microbiol. 22 (5), 815-826 (1996)
MEDLINE        97126810
REFERENCE      2 (bases 1 to 2374)
AUTHORS        Vrijic,M.M.
TITLE          Direct Submission
SUBMITTED      (07-MAR-1996) M.M. Vrijic, Institut fuer Biotechnologie
JOURNAL        1, Forschungszentrum Juelich, Postfach 1913, D-52425 Juelich, FRG
FEATURES
source
1..2374
/organism="Corynebacterium glutamicum"
/strain="R127"
/db_xref="taxon:1718"
complement(82..954)
/gene="lysG"
complement(82..954)
/gene="lysG"
/notes="Lysine governor"
/codon_start=1
/transl_table=11
/product="Lysine export regulator protein"
/protein_id="CAA65323.1"
/db_xref="GI:1729754"
/translation="MNPIDLTLLSIDESFEGASLALSISPSAVSQRVKALEHHVG
RVLSVTPQAKAEAGEVILVQAAKMWLLQAEIKAOISGRLAIEPILTIAINADSLSTW
FPVFEVNSWGATLTLEDEAHTLSLLRGDVLGAVTREANPVAGCEVVELGTMR
HLAIFPSLRDAYWDGKLDWAAMPVLRGPKDYLDQDLRDLDGVDGVPVGRRRYSVPS
AEGFGAIRGLGWGLLPETQAAAPMLKAGEVILLDEIPIDTPMIWQRWRLSRLARL
TDADVDAATIEGLRP"
1025..1726
/gene="lysE"
1025..1726
/gene="lysE"
/codon_start=1
/transl_table=11
/product="Lysine exporter protein"
/protein_id="CAA65324.2"
/db_xref="GI:13397387"
/translation="MEFITGLLLGASLLISICPQNVLYIKQIKREGLIALLVGLI
SDVFLFIAGTLVDLSUNAPIVLDIMRWGGIAYLLWFVMAAKDAMTNKVEAPQIE
ETEPTVPDTPDGGSVAVDTNRVRVEVSVDKORVWKPMLMAIYLTLNPNAYLDA
FVFIGVGAQYGDTRWIFAAAGAFASLIWFPLVGFCAALSRPLSPKVVRRINVVV
AVVMTALAIKMLMG"
complement(1723..2352)
/gene="orf3"
complement(1723..2352)
/gene="orf3"
/codon_start=1
/transl_table=11
/protein_id="CAA65325.1"
/db_xref="GI:1729756"
/db_xref="GI:1729756"
/translation="MYALRDIASGKALYVIGSIYPGELTAEEAFMAEEGCPILLIHQ
PSYSLINRVEPDPDGLNQAANNGLVIAFSPLOAGLLTDKYLIDGIPGSRASQ
GKSLSGMLNNVNDMVRKINDIAOERGOSLAQMALAWLREOGEYGAQDTVTSALIGA
SSVEQLDLSLNNLFEFDAELEAIDEISHDAGINIAKATDSKRTEN"

```

```

BASE COUNT      526 a      640 c      648 g      560 t
ORIGIN
Query Match      100.0%; Score 873; DB 2: Length 2374;
Best Local Similarity 100.0%; Pred. No. 1,le-235;
Matches 873; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 atgaacccattcaactggacacatttgcctcaatcattgatgaagcagcttcgaaggc 60
|||||
Db 954 ATGAACCCCAATTCAACTGGACACTTTGCTCTCAATCATGTATGAAGCAGCTTGAAGGC 895
Qy 61 gctcccttagccctttccatttcccttcggggtgagtcagcgcttaagtctctcag 120
|||||
Db 894 GCCCTCTTAGCCCTTTCCATTTCCCTCGCGGTGAGTCAGCCGCTTAAAGCTCTCGAG 835
Qy 121 catcagctgggtcagtggttggtatcgccaccacccgcccgaagcaaacgagcggt 180
|||||
Db 834 CATCAGCTGGGTGAGTGTGGTATCGCGCACCCCAACCGGCCCAAGCAACGGAAGCGGGT 775
Qy 181 gaagtccttgtcaagcagcgcggaataatggtgttgcgtcaagcagaactaaagcgcaa 240
|||||
Db 774 GAAGTCCTTGTGCAAGCAGCGCGGAAAAATGCTGTGCTGCAAGCAGAACTAAAGCGCAA 715
Qy 241 ctatctggagcgcttgcgtgaatacccgcttaaccatcccatcaacgagagctctatcc 300
|||||
Db 714 CTATCTGGACGCTTGTCTGAAATCCCGTTAACCATGCCCATCAACGGCAGATTCTCTATCC 655
Qy 301 acatggttctccctcggttcaacgaggtagcttcttggggtgagcagcgctcaecgtg 360
|||||
Db 654 ACATGGTTTCTCCGCTGTTCAACGAGGTAGCTTCTTTGGGGTGGGCAACGCTCACGCTG 595
Qy 361 cgcttgaagatgaagcgcacacattatccttgcctgcggcggtgagatgttttaggagcg 420
|||||
Db 594 CGCTTGAAGATGAAGCGCACACATTATCTCTGCTGCGCGGTGGAGATGTTTTAGGAGCG 535
Qy 421 gtaaccgctgaagctaataccgctgagcggtatgtaaatagtagaactggaacatcgcg 480
|||||
Db 534 GTAAACCGTGAAGCTAATCCGCTGGCGGGATGTGAAGTAGTAGAATCTGGAACCATGCGC 475
Qy 481 cacttgccattgcaacccctcattcgggatgcctcacatggttgatgggaaactagat 540
|||||
Db 474 CACTTGCCCATTGCAACCCCTCATTTGCGGGATGCCCTACATGTTGATGGAAACTAGAT 415
Qy 541 tgggctcgatgcccgtcttaccgcttcggttcggttcccaaaagatgtgcttcaagaccgtgacctg 600
|||||
Db 414 TGGCTCGGATGCCGCTCTTACGCTTCGGTCCCAAGAGATGCTTCAAGACCGGTGACCTG 355
Qy 601 gacggcgcgctcgatgctcgttgggcgagcgcgctatccattgctccgctggcggaag 660
|||||
Db 354 GACGGCGCGCTCGATGCTGCTGTGGGCGCGAGCGCGGTATCCATTTGCCGTCGGCGGAA 295
Qy 661 ggttttggtgagcgaattccgagggccttgggtgggagcttcttcccgaaacccaagct 720
|||||
Db 294 GGTTTTGTGAGGCAATTCGCCGAGGCGCTTGGTTGGGGACTTCTTCCCGAAACCAAGCT 235
Qy 721 gctccatgctaaagcagagagatgctcctcctcgatgagatacccatgacacaccg 780
|||||
Db 234 GCTCCCATGTAAAGCAGAGAGAGTATCCTCCTCGATGAGATACCCTATGTACACACCG 175
Qy 781 atgtattggcaacgatggcgcttgggaatctagatctcttagtaactcacagacgcgtc 840
|||||
Db 174 ATGTATTGGCAACGATGGCGCTGGGAATCTAGATCTCTAGCTAGACTACAGAGCGCGTC 115
Qy 841 gttgatcagaatcagaggattgcgcccttag 873
|||||
Db 114 GTTGATCAGCAATCGAGGGATTGCGGCCTTAG 82

```

RESULT 3
A93933/c
LOCUS
DEFINITION

A93933 2374 bp DNA PAT 22-JAN-2000
Sequence 2 from Patent WO9723597.

ACCESSION A93933
VERSION A93933.1 GI:6742037
KEYWORDS
SOURCE Corynebacterium glutamicum.
ORGANISM Corynebacterium glutamicum.
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Corynebacterineae; Corynebacteriaceae;
Corynebacterium.
REFERENCE 1 (bases 1 to 2374)
AUTHORS Vrljic.M. and Eggeiling.L.
TITLE PROCESS FOR THE MICROBIAL PRODUCTION OF AMINO ACIDS BY BOOSTED
ACTIVITY OF EXPORT CARRIERS
JOURNAL Patent: WO 9723597-A 2 03-JUL-1997;
KERNFORSCHUNGSANLAGE JUELICH (DE); VRLJIC MARINA (DE)
FEATURES
Source 1..2374
/organism="Corynebacterium glutamicum"
/db_xref="taxon:1718"
BASE COUNT 526 a 640 c 648 g 560 t
ORIGIN

Query Match 100.0%; Score 873; DB 9; Length 2374;
Best Local Similarity 100.0%; Pred. No. 1.1e-235;
Matches 873; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atgaacccattcaactgacacattgtctcaatcattgatgaagcagcttcgaagc 60
|||||
DB 954 ATGAACCCATCACTGACACTTGTCTCAATCATTTGATGAAGCAGCTTCGAGGC 895
|||||
QY 61 gctcttagcctttccatttccctcctggcggtgagtcagcggttaaaactctcgag 120
|||||
DB 894 GCTCTTAGCCTTTCCATTTCCCTCCCTCGCGGTGAGTCAGCGCTTAAAGCTCTCGAG 835
|||||
QY 121 catcagtggtgagtggtgtatcgccaccaccgcaaaagcgaagcgaagcggt 180
|||||
DB 834 CATCACTGGGTGCGAGTGTGTATCGCGCACCCCAACCGGCCCAACCGCGGT 775
|||||
QY 181 gaagtccttgcgaagcagcggaataatggtgtctcaagcagcaaaactaaagcgcaa 240
|||||
DB 774 GAAGTCTTGTGCAACGACGCGGGAATATGTTGCTGCAGCAGCAAACTAAAGCGCAA 715
|||||
QY 241 ctatctggagccttgcgtgaatcccggttaacacatgccatcaacgcagatcgctatcc 300
|||||
DB 714 CTATCTGGAGCGCTTCTGCTGAAATCCGTTAAACCATCGCCATCAACGCATTCGCTATCC 655
|||||
QY 301 acatggtttccctccgtgttcaacgaggtagcttcttgggggtggagcaagcgtcacgctg 360
|||||
DB 654 ACATGTTTTCCTCCCGTGTTCACAGGAGGTAGCTTCTTGGGGTGGAGCAACGCTCAGCGTG 595
|||||
QY 361 cgcttggagatgaagcgcacacattatcttgcctgcggtggagatgttttagagcg 420
|||||
DB 594 CGCTTGGAGATGAAGCGCACACATATCTTCTGCTGCGCGGTGGAGATGTTTGAAGGCG 535
|||||
QY 421 gtaacccgtgaagcctaactccctggcggtgagtagtagtaagtagtaagccttggaaaccatggcg 480
|||||
DB 534 GTAACCCGTGAAGCTAATCCGCTGGCGGATGTGAAGTAGTAGAACTTGAACCATCGCGC 475
|||||
QY 481 caattggccattgcaacccctcattgcgagatgctacatggttgatggaaactaagt 540
|||||
DB 474 CACTTGGCCATTGCAACCCCTCATTTGCCGGATGCCCTACATGTTGATGGGAACATAGAT 415
|||||
QY 541 tgggctgcgagtcgccgttttacgcttcgctcccaagaatgcttcaagacccgtgacctg 600
|||||
DB 414 TGGGCTGCGATGCCGCTCTTACGCTTCGCTGCCAAGATGTGCTTCAAGACCGTGAACCTG 355
|||||
QY 601 gacggcgcgctcgatggtccttggggcgagcgagcgatccattgtcccgctcgcgga 660
|||||
DB 354 GACGGCGCGCTCGATGTCCTTGGGGCGCAGCGCGGTATTCATTGTCCTCGCGCGGAA 295
|||||
QY 661 ggttttggtaggcaattccgcagcgcttgggtgggactcttcccgaaaccccaagct 720
|||||
DB 294 GGTTTTGTGAGGCAATTCGCCGAGCGCTTGGTGGGAGCTTCTTCCCGAAACCAAGCT 235
|||||

QY 721 gctcccatgctaaaagcaggaagatgacctcctcgatgagatccccattgacacacg 780
|||||
DB 234 GCTCCCATGCTAAAAGCAGGAGAGTGATCTCTCGATGAGATACCCATTGACACACG 175
|||||
QY 781 atgattggcaacgatggcgctggaatagatctctagctagctacacagccgctc 840
|||||
DB 174 ATGATTGGCAACGATGGCGCTGGAATAGATCTCTAGCTAGACTACAGACGCGCTC 115
|||||
QY 841 gttgatgcagcaatcgaggagattgcgacctag 873
|||||
DB 114 GTTGATGCAGCAATCGAGGATTGCGGCGCTTAG 82
|||||

RESULT 4
SC5F8/c SC5F8 33285 bp DNA BCT 09-JUN-2000
LOCUS Streptomyces coelicolor cosmid 5F8.
ACCESSION AL357613
VERSION AL357613.1 GI:8347023
KEYWORDS amino acid decarboxylase; anti-sigma factor; integral membrane protein; LuxR-family regulator; LysR-family transcriptional regulator; membrane transport protein; oxidoreductase; regulatory protein; RNA polymerase sigma factor; serine/threonine protein kinase; stress-inducible protein; TetR-family transcriptional regulator; thioredoxin reductase; threonine dehydratase; two-component sensor histidine kinase.
SOURCE Streptomyces coelicolor A3(2).
ORGANISM Streptomyces coelicolor A3(2).
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Streptomyceinae; Streptomycetaceae; Streptomyces.
REFERENCE 1 (bases 1 to 33285)
AUTHORS Redenbach, M., Kieser, H.M., Denapalte, D., Eichner, A., Cullum, J., Kinashi, H. and Hopwood, D.A.
TITLE A set of ordered cosmids and a detailed genetic and physical map for the 8 Mb Streptomyces coelicolor A3(2) chromosome
Mol. Microbiol. 21 (1), 77-96 (1996)
97000351
JOURNAL 2 (bases 1 to 33285)
MEDLINE Seeger, K.J. and Harris, D.
REFERENCE 3 (bases 1 to 33285)
JOURNAL Thompson, N.R., Parkhill, J., Barrell, B.G. and Rajandream, M.A.
AUTHORS Submitted (05-JUN-2000) Streptomyces coelicolor sequencing project, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
TITLE CB10 ISA E-mail: barrell@sanger.ac.uk Cosmids supplied by Prof. David A. Hopwood, [3] John Innes Centre, Norwich Research Park, Colney, Norwich, Norfolk NR4 7UH, UK
JOURNAL
COMMENT Notes:
Streptomyces coelicolor sequencing at The Sanger Centre is funded by the BBSRC and Beowulf Genomics
Details of S. coelicolor sequencing at the Sanger Centre are available on the World Wide Web.
(URL: http://www.sanger.ac.uk/projects/S_coelicolor/)
CDS are numbered using the following system eg SC7B7.01c. SC (S. coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary strand).
The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous.
The length in codons is given for each CDS.
Usually the highest scoring match found by fasta -o is given for CDS which show significant similarity to other CDS in the database. The position of possible ribosome binding site sequences are given where these have been used to deduce the initiation codon.
Gene prediction is based on positional base preference in codons using a specially developed Hidden Markov Model (Krogh et al., Nucleic Acids Research, 22(22):4768-4778(1994)) and the Frameplot program of Bibb et al., Gene 30:157-66(1984) as implemented at <http://www.nih.gov.jp/jun/cgi-bin/frameplot.pl>. CAUTION: We may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtg, ttg or (att)) which is preceded by an upstream

ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.

IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions.

Cosmid 5F8 lies on genomic restriction fragment AseI-A bordered by cosmids 5H1 and 4G10.

FEATURES

source	1. .33285 /organism="Streptomyces coelicolor A3(2)" /strain="A3(2)" /db_xref="taxon:100226" /clone="cosmid 5F8" complement(1. .344) /gene="SC5F8.01c" complement(1. .108) /note="nominal overlap with cosmid SC5H1 between bases 1. .108."
misc_feature	/gene="SC5F8.01c"
CDS	complement(<1. .344) /gene="SC5F8.01c" /note="SC5F8.01c, possible serine/threonine protein kinase, partials CDS, len> 114 aa. Similar to many proposed Streptomyces coelicolor serine/threonine protein kinases e.g. TR:CAB82014 (EMBL:AL161755) SCD63.07 (717 aa), fasta scores opt: 222.0 E(): 6.8e-05 33.0% identity in 103 aa overlap. Overlaps and extends into CDS SC5H1.01 on the adjoining cosmid."
gene	/codon_start=1 /transl_table=11 /product="putative serine/threonine protein kinase (fragment)" /protein_id="CAB93729.1" /db_xref="GI:8347024"
CDS	/translation="MGRHVTHLVAGRYRLFEVQVRETRNCVSGEDATGTRGPLV VRLGDLDRGA" complement(492. .1451) /gene="SC5F8.02c" complement(492. .1451) /note="SC5F8.02c, possible threonine dehydratase, len: 319 aa. Similar to many including: Escherichia coli SW:THD2_ECOLI (EMBL:M21312) threonine dehydratase, catalytic (EC 4.2.1.16) TdcB (329 aa), fasta scores opt: 609 z-score: 676.6 E(): 3.2e-30 39.4% identity in 312 aa overlap. Contains a Prosite hit to Serine/threonine dehydratases pyridoxal-phosphate attachment site and a Pfam match to entry PF00291 S_T_dehydratase, Pyridoxal-phosphate dependent enzyme."
misc_feature	/codon_start=1 /transl_table=11 /product="putative threonine dehydratase." /protein_id="CAB93730.1" /db_xref="GI:8347025"
misc_feature	/translation="MTAITEIAAERIAHVVRTPSPGLSALLGVPTAKLELL QRTGSPKAGATAKLLSLEAEAGVAVSGNGHGIADVMAALDVKATVVMPTA PARSVIAEAGALVRLTDGMSAFALVRLREGILVHPFDDPVVAGQGVGLEF AEDASDLTVSISGGGLGSAFALRALRPGVRVGVWETEGAEAMSRALAGGPT VLSSVVTLSAPSVSRSLTVHVAELVTVLPVVDREAVRSGLSLAIEAHKVVTEPAAG CLLPAAVRVYRGDARGIGLVVCGGNATVGDMAVWADRFLR" complement(546. .1412) /gene="SC5F8.02c"
gene	/note="Pfam match to entry PF00291 S_T_dehydratase, Pyridoxal-phosphate dependent enzyme, score 210.10, E-value 3.3e-59" complement(1287. .1328) /gene="SC5F8.02c" /note="PS00165 Serine/threonine dehydratases pyridoxal-phosphate attachment site" complement(1477. .2154)

CDS	/gene="SC5F8.03c" complement(1477. .2154) /gene="SC5F8.03c" /note="SC5F8.03c, unknown, len: 225 aa. Similar to several proteins of undefined function including: Deinococcus radiodurans TR:Q9RWP1 (EMBL:AE001920) conserved hypothetical protein (254 aa), fasta scores opt: 467 z-score: 555.6 E(): 1.8e-23 36.6% identity in 205 aa overlap and Streptomyces coelicolor TR:CAB70638 (EMBL:AL137242) SC8F4.12C (239 aa), fasta scores opt: 397 z-score: 474.2 E(): 6.1e-19 36.1% identity in 183 aa overlap." /codon_start=1 /transl_table=11 /product="hypothetical protein SC5F8.03c." /protein_id="CAB93731.1" /db_xref="GI:8347026"
gene	/translation="MKGDLFSSEHMVQATPCMTVNSKCIYAVNGEMLAROGAMI AYKGNLQFERKGGVCGMLKRAVTEGGLPLMAVGGGEAFANEAOCFFVEVPEQDE FTVNGRNVLFCFASLRIATVKGSGAGGLFNFTVGTGRLGLVCEGNPLVPSVQ QYPVHVDTDAVVGWSAGLATSLHRSQSIGSMRGSGEAVQLVLOGEGFVVVRPSEAT PQKPPQOH"
CDS	complement(2166. .3470) /gene="SC5F8.04c" complement(2166. .3470) /gene="SC5F8.04c" /note="SC5F8.04c, unknown, len: 434 aa. Similar to Streptomyces coelicolor TR:Q9XW9 (EMBL:AL096743) putative transport associated protein, SC17.21 (399 aa), fasta scores opt: 428 z-score: 449.2 E(): 1.5e-17 38.3% identity in 209 aa overlap." /codon_start=1 /transl_table=11 /product="hypothetical protein SC5F8.04c." /protein_id="CAB93732.1" /db_xref="GI:8347027"
gene	/translation="MSDLVPGGNVPLPGGVSVPVPGGFDVSAVLTDEGKVGGDADF VFYNQEPAGRLRDLDTLVDFARLRGAARVTVAVGSDPTGGLALPSPVTLVTD RGRTVAFTPARGRETVLLVLFYRRGGKVKVVALGQYADGLAGLDFGVEVTD APPEPATPEPLTPGADSDGFLNLSNARAAGSPAVRPDPLRSARAAHAAAMAAAG TLSIETRDGVSVHVVVSAFVLTVEHLVSGPRTPAEFVAYCLRAERTRLTHDTA FTHAGWACVGGPSGDTWTALWAVPLTPGLARTTAEVVGLTNREACAGLPAVLD ARLTAQAHSADMTVDYSHDTPGGAPDRAAAGADRRSNGENIACGQSPRAEV VEGWNMSPGRANILEAGTHGVGLAGGRAGTYWTQLGG"
CDS	complement(3562. .6528) complement(3562. .6528) /note="SC5F8.05c, possible LuxR-family regulator, len: 988 aa. Similar to several other putative regulators from Streptomyces coelicolor including: TR:Q9XAN4 (EMBL:AL079355) SC4C6.06 (893 aa), fasta scores opt: 305 z-score: 328.7 E(): 7.7e-11 26.4% identity in 910 aa overlap and TR:Q92573 (EMBL:AL035569) SC8D9.18 (1091 aa), fasta scores opt: 857 z-score: 926.1 E(): 0.36.5% identity in 1077 aa overlap. Contains a Prosite hit to PS00017 ATP/GTP-binding site motif A (P-loop) Pfam match to entry PF00196 GerE, Bacterial regulatory proteins, luxR family."
misc_feature	/codon_start=1 /transl_table=11 /product="putative LuxR-family regulator." /protein_id="CAB93733.1" /db_xref="GI:8347028"
misc_feature	/translation="MGSVPQTSITSLPVGREDELAELTCVLERARAGEARVAVAGDA GVGKRTLHEAAGRAAAAGTTLTGCHVDLGVGLPYLFFTEILGVLADEAFNAVLA GHPVADRLGGPDGDDAAPSRLRLEFEGVAALLTELADVAPLLVLEDLHWADQSSR DLIRFLGVLQPGAGAPRLALFASYRADDDCHRRHPLRLPLAEVLRPLGVPERLE LRPLPDSVARLVRLRRRLRRTVTRIVERAEGNAFYAEELVATAPAGVPSGL ADVLIRFEQLSETAQOVLRTAAVAGRRVGHDLRLDRAVGLPEELESALREAVERQLL

Query Match 16.1%; Score 140.6; DB 3; Length 33285;
Best Local Similarity 50.9%; Pred. No. 6.6e-29;

[illegible]

Qy	724	ccatcgtctaaacagcagagaagtgatctctcctcctcgatgagatcaccatgatgcacacccgatg	783
Db	5964	TCCTCCGGTTCCCGCATGGATCGGTTCTCGTACGGGTCTGGGACATACACTCGAGCTGCCCTCTC	6023
Qy	784	tattggcaacgatggcgccctgggaatctagatctctagcttagactcacagacgcgcgtgctt	843
Db	6024	TATTGGCAATGCTTGGAACTGGACAGTCCGATCATCGCGCGAATTACCGACACGCGTGAGG	6083
Qy	844	gatgcagcaatcg	856
Db	6084	CGCGCGCAAGCG	6096
RESULT	7		
LOCUS	MTCX39	38500 bp	DNA
DEFINITION	Mycobacterium tuberculosis H37Rv	complete genome;	segment 89/16
ACCESSION	274025	AL123456	
VERSION	274025.1	GI:3261586	
KEYWORDS			
SOURCE	Mycobacterium tuberculosis.		
ORGANISM	Mycobacterium tuberculosis		
	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;		
	Actinomycetales; Corynebacterineae; Mycobacteriaceae;		
	Mycobacterium; Mycobacterium tuberculosis complex.		
REFERENCE	1 (bases 1 to 38500)		
AUTHORS	Cole, S. T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., Gordon, S. V., Eiglmeier, K., Gas, S., Barry III, C. E.,		


```
CDS
/ gene="Rv1984c"
/ complement(4656..5309)
/ note="Rv1984c, (MTCY39.35), len: 217. Probable cutinase precursor with N-terminal signal sequence, similarity to CUT1_AUTBR P41744 cutinase precursor, (209 aa) FASTA scores: opt: 283, z-score: 322.0, E(): 2.2e-11, (32.6% identity in 193 aa overlap). Also similar to M. tuberculosis Rv3452, Rv3451, Rv2301, Rv1758, Rv3724."
/ codon_start=1
/ transi_table=11
/ product="hypothetical protein Rv1984c"
/ protein_id="CAA98399.1"
/ db_xref="GI:1403471"
/ transi_table="MTPRSLVRIYGVVVVATTLALVSPAGRAAHADPCSDIAVVFAR GTHOASGLGDVEAFVDSLTQVGGVAVVNPASDDYRASASNGSDASAHIOR TVASCPNTRVLGYSOGATVIDLSTAMPAPVADHVAALFGEPSGSSMLWGGG SLPTIGLYSSKTNLCAPDPICTGGNIMAHVSYVQSGMTSQATFAANRLDHAG" complement(4950..4967)
/ gene="Rv1984c"
/ note="PS00155 Cutinase, serine active site"
/ complement(5739..6650)
/ gene="Rv1985c"
/ complement(5739..6650)
/ gene="Rv1985c"
/ note="Rv1985c, (MTCY39.34), len: 303, LysR family member similar to many regulatory proteins, especially ICIA_ECOLI P24194, chromosome initiation inhibitor, (297 aa), and P94632 LYSINE EXPORT REGULATOR PROTEIN. (290 aa). FASTA scores, P94632 LYSINE EXPORT REGULATOR PROTEIN opt: 705 z-score: 850.6 E(): 0; 42.7% identity in 288 aa overlap; and ICIA_ECOLI P24194 opt: 520, z-score: 632.5, E(): 1.1e-28, (35.8% identity in 285 aa overlap); contains PS00044 Bacterial regulatory proteins, lysR family signature. Also contains helix-turn-helix motif at aa 22-43, (+5.52 SD)."
/ codon_start=1
/ transi_table=11
/ product="hypothetical protein Rv1985c"
/ protein_id="CAA98410.1"

Query Match 13.9%; Score 121; DB 3; Length 38500;
Best Local Similarity 49.8%; Pred. No. 2.2e-23;
Matches 425; Conservative 0; Mismatches 410; Indels 18; Gaps 4;

Qy 13 caactggacatttgctcoaatcattgatgaagcagcttcgaagcgccctccttagcc 72
Db 6623 CAGCTGGCCGCGATGGTCGCGTGTGAACTGGCAGCTTCGATGCGCCGCGGAGCGC 6564

Qy 73 ctttccattccccctcgccggtgagtcagcgcgcttaagctctcgagcatcacgtgggt 132
Db 6563 CTACATGTCACCCGCTCGCTGTCAGTCAGCGCATCAAGTCTGTGGAGCAGCAGTGGC 6504

Qy 133 cgagtggttgatcgccaccaccacccggccaaagcaacccgagcggtgagtccttctg 192
Db 6503 CAGGTGCTGGTGGTCAGGGAAGAACCATGTGCGGCGCAGCACCGAGGTATCCCGCTGTG 6444

Qy 193 caagcagcgcggaataatggtttgctgcaagcagaactaaagcgaactatctgg--ac 250
Db 6443 CGGTGGCCGCGCAACAGACGCTGCTCGAGTCGCGAGGGCTCGCTGAATGGGTGGCAAC 6384

Qy 251 gcttctgtgaa----atcccggttaaccatgccatcacgagattcgctatccacatgg 306
Db 6383 CGCTCGTGAACACGACGCGGATCACCATTGCGGTAAACGCGCATTCATCGCGGACATGG 6324

Qy 307 ttctctcccggtgtcaacgaggtagctcttctgggtggagcaacgctcacgctgcgttg 366
Db 6323 TTTTCGCGCGGTTCGACG-----GTCTCGGCGACGCTCTGCTCGAGCTTCGGATC 6273

Qy 367 gaagatgaagcgacacattatctctgctgcgcgctggagatgttttaggagcgctaac 426
Db 6272 GAGGACGAGGACCATTCGCGCGCGCTGCTACGGAGGGGTGTGGCGATGGGCGCGGTACC 6213
```

```
Qy 427 cgtgaagctaatacccgctggcggtggtgaagttagtagaacttggaacacatcgccacttg 486
Db 6212 ACCGAGCGGAACCCGCTGCCGGGTGCGCCGGTGCACCCGCTGGTGAAATCGCTACCTA 6153

Qy 487 gccattgcaacccctcattgcgggattccctacattggttgatgggaaactagattggct 546
Db 6152 CCAGTGGCCAGCAGGCCCATTCGTCCAGCGGCATCTATCCGACGGGTTCACCTGCCGCCGCG 6093

Qy 547 gcgatgcgcgtcttacgcttcacgcttcgccaaagatgcttcaagacgcgtgacctgacggg 606
Db 6092 GCGGCTAAAGCTCCGTCACTGCGGTGGAATCGTGACGATGGCTGCAGGACATGTTGCTG 6033

Qy 607 cg---cgtcgatgcttcgtgtggcgcgagggcgctgctcattgctccctcgcgggaaggt 663
Db 6032 CGTAAGGCTTTCGTGCGGCCATCACCAGACCGACGACATTTGTGCCGACACAGAGGC 5973

Qy 664 ttggtgaggaattcgcgagggcctgtgtggtgggaactcttccgaaacccaagctct 723
Db 5972 TTCACCGCGGACGCGCGCGGCTGGGATGGGATGGGCGATGTTCCCGGGAAGCTGGCAGCA 5913

Qy 724 cccatgctaaagcaggaagatgctcctcgatgagatgacccattgacacacccgatg 783
Db 5912 TCTCGCTTGGCGATGGATCGTTCGACGGGTCTGCGACATACACCTCGAGTCCCTCTC 5853

Qy 784 tattggcaacgatggcgctggaatcttagatctcttagcttagctacacagacgcgtgctt 843
Db 5852 TATTGGCAATGCTGGAACCTGGACAGTCCGATCCGCGCAATATTACCGACACGGTGAGG 5793

Qy 844 gatcgagcaatcg 856
Db 5792 GCGGCGGCAAGCG 5780

RESULT 8
AB011413/c AB011413 12070 bp DNA BCT 07-AUG-1998
LOCUS Streptomyces griseus genes for Orf2, Orf3, Orf4, Orf5, AfsA, Orf8, partial and complete cds.
DEFINITION AB011413
ACCESSION AB011413 1 GI:3401946
VERSION Orf8; AfsA; Orf5; Orf4; Orf3; Orf2.
KEYWORDS Streptomyces griseus DNA.
SOURCE Streptomyces griseus
ORGANISM Actinomycetales; Streptomycineae; Actinobacteria; Actinobacteridae; Streptomyces.
REFERENCE 1 (bases 1 to 12070)
AUTHORS Uneyama,T.
TITLE Direct Submission
JOURNAL Submitted (23-FEB-1998) to the DDBJ/EMBL/GenBank databases. Takashi Uneyama, University of Tokyo, Department of Agriculture and Life Sciences, Yayoi 1-1-1, Bunkyo-ku, Tokyo 113, Japan (E-mail:aa67103@hongo.ecc.u-tokyo.ac.jp, Tel:+81-3-3812-2111)
2 (sites)
REFERENCE Uneyama,T.
AUTHORS Open reading frame encoded around afsA gene
TITLE Unpublished (1998)
JOURNAL Location/Qualifiers
FEATURES
source 1..12070
/organism="Streptomyces griseus"
/db_xref="taxon:1911"
/complement(2660..3511)
/gene="orf2"
/complement(2660..3511)
/gene="orf2"
/function="sensor histidine kinase"
/codon_start=1
/transi_table=11
/product="Orf2"
/protein_id="BAA32130.1"
/db_xref="GI:3401947"
/transi_table="VAAPVPAHPGPEAAAPAAARLTAAATLTITVVGLACLSWLVIIR TDDRARMDAEYDEMRRRASVAASLVYVEDDIRSAGPAEHRSLVLSLVIGAGALTALA
```

ALTGHVLSRSGCPANWALEQQERLEADAAHELRTVPVYMRGSVEVAAGAPGLEGOL
PRIRRAADMAVDVNNLTLRGRLEAGRTSYGRSLRLDLQDLVEVCAELPEGGPGLER
LEESVVAADAAVLVNRNLLNAVHRHGPGRGAAGRGRTAGDRVRTGVRVADGPVAV
APARLPS"

CDS

complement(3523..>4197)

/function="Two-component regulator protein"

/codon_start=1

/transl_table=11

/product="Orf3"

/protein_id="BAA321131.1"

/db_xref="GI:3401948"

/translation="MRVLVEVDDDLRDVIGAGLRDGGFAMDCASDWPADVLLHLLSA
YDCVLDVMPVSGDTAPLEGRRAGSVFVLCLTALDSLDERLGRLESGADYLAKP
FMRELVLVRGLSRASRLPSFLCADVMDVARHEVRGGLVLSLSPKRYAVLQO
LLVHRTVTRTGLLEHCWDEMDPVSNVVDVAVAGLRRLKGLSPGLVHTVRQGFLLS
ADPGPS"

CDS

complement(4333..>5250)

/function="methyltransferase"

/codon_start=1

/transl_table=11

/product="Orf4"

/protein_id="BAA321132.1"

/db_xref="GI:3401949"

/translation="NATQALGTALRLGVFDRIGTCELRAELAGSLGTHPRATLRLR
ALAGLQLLSEPPGAFRTTAAGNVLRADAPGTHVAMARFTDFVMLRGWDLDESVRT
GETTFDTGTFGLRHEHPSELAFNEAMSGQRLTAETVPHHYDFGRFORLVDIG
GGDTLLASILRAHOBPRGVLEDTAEGLAQAAPRLAREGLDGRVLTLETGDFASAPAG
GDLVLSNLSIHWDDAVRGLRHIRDVLPDHSLLIVELVLPATVPADRPDNNVLSDL
NMLVNVGGRERTADDPAAALCTAGFACGA"

CDS

complement(5450..6361)

/function="regulator protein"

/codon_start=1

/transl_table=11

/product="Orf5"

/protein_id="BAA321133.1"

/db_xref="GI:3401950"

/translation="MSELPDQVRLVAVVDEGTSDSANTALRLTPSAVSQRVKALEQ
RTGRVLLMTKTPVRPTESGEVVALQARLEHQAALGMSGPGPEPTLLPIAVNAAQ
LATWQPCGACRWDSSATTAPGGPGPYRPAAGAGDGRDLVAGGSDGLFGFPAAG
RMRYLPVAPSGFADRWLRDRGTALRELIGEAPVVCVDRRDDLDQAFVRLGPGARPS
ARRHLVPTSEGAFANAVASGMGMVPEVQAEPLSDGRULVRLAPEPTVDVPLYWQWQK
LESPLAAVAEAAVEAAEALDEAPPN"

CDS

7486..8391

/codon_start=1

/transl_table=11

/product="AfsA"

/protein_id="BAA321134.1"

/db_xref="GI:3401951"

/translation="MDAAEVVHPVGIEMVHRTREPDAFPNNVRLGRDRFAVEAVLP
HDPFPAPVGDLDHDLPLVAEMRQAAMLAHFAGYIPIGLYHFLITELDYVCHPEHLG
VGGEPTGELVFCSDLKWRAGLPAQGRVGMVHRGDLRAATGVAATRFSTPKAYRM
RGDVPVEGLSPETAPVPASPAGRARVEDVLSGTGREGVWELRVDTRHPTLFRPND
HYVGMILLAAARQAACLAVAGPAGIIPVEARTFRHYSERGSPCWIGAVVQCADEDTV
TVRVTHGDGEVTFSTVLSGPRANG"

CDS

10558..11598

/function="alcohol dehydrogenase"

/codon_start=1

/transl_table=11

/product="Orf8"

/protein_id="BAA321135.1"

/db_xref="GI:3401952"

/translation="MTTVAAYAPRAKAPLERTTIERRPVGEFDILIDIKFAGICHSD
IHQANDGEGIFPVPVGHETIAGIVAETPGVTRKVGDRVGVGMVDSGCTCDACLM
GREGHACBNTQTYNALDRNGEPTYGYSTHLVTFKFAISPEGLADEAAPLILCAG
ITTPSPLRWAGPGKVAHVGLGLGHMAVKIAHALGAENVLSQSLRKKDDGLKLG
ADHYATGDPRTFEELAGTFDVLSTVSAPLDFGAYLALLRTEGLVNVGAPEPVSL
NLFSILGNRSIAGSAIGIETQEMLDCAVHGLCAELIEVIGACQVNEATVERVLASD
VRYRVTDITAT"

BASE COUNT
ORIGIN

1647 a 4478 c 4272 g 1673 t

Query Match

12.4%; Score 108.6; DB 1; Length 12070;

Best Local Similarity 48.8%; Pred. No. 7.7e-20;
Matches 425; Conservative 0; Mismatches 424; Indels 22; Gaps 4;

QY 13 caactgagacacttctctcaatcattgatgaaggcagcttcaaggcgcctcttagcc 72

Db 6340 CAGGTGCGGACCTTGTGGCGGTGGTGCAGCAAGAGCACCTTCGACTCGCGGGGACCGCC 6281

QY 73 ctttcatttccccctcgcgcgtgagtcagcggttaaaagctctcagacatcagcglgggt 132

Db 6280 CTGCGGCTGACCCCGTTCGGCGGTTCATGTCACGGGTGAAGCGGTGGAGCAGCGACCGGC 6221

QY 133 cgagtgttggtatcgcgcaccccaacggccaaagcaacacggcgaagcgggtgaagtccttgtg 192

Db 6220 CGGTACTGCTGATGCGCAGCAAGACGGGTGCGGCCACCGAGTCCGGTAGGTTGGTGGCG 6161

QY 193 caagc---agcgcggaataatggtgtgtcgcgaagcagaactaaagcgaactatctgga 249

Db 6160 CTCGCCCGCCAGCTGGCCAGGCTGGAGCAGAGGCGCAGGCGCTGGGATGTCCGGG 6101

QY 250 cgcttctgtaaacctcccttaaccatcgccatcaacgcagattcgctatccacatggttt 309

Db 6100 CCGGCGAGCGGACGCTCTCTGCGGATCGCGGTGAACCTCCGATTCCTTGGCGACCTGG--T 6043

QY 310 cctcccggtttcaacgaggtagcttcttgg99tgg9gcaacgctcagcgtcgcttgaa 369

Db 6042 CTTGGCAGCCCTCGCGCGGTGCGCGAGGAGTTGGGACTCTGCTAGCACTGCGCCGGGAG 5983

QY 370 gatgaagcgcacattatcttctgctgcgcgtggagatgttttagagcggtaaacccgt 429

Db 5982 GACCAAGACCATACGGCGCCCTCTGTCGGGAGGGGTGCGGATGGCCGGCGGTGACCTCG 5923

QY 430 gaagctaateccgtggcggtgtagtagtagtaagtagtaagacttgaacacttgcc 489

Db 5922 TCGCGGAGGCGGTGACGGGCTTTCGGTCCGGCGGTGGCGCATGCGCTACTTGCCTGG 5863

QY 490 attgcaacccctcattggcggtgctacatagttgatgggaaactagattgggtgctg 549

Db 5862 TGCCCACTCCCGCTTTTCGCCGACCGGTGGGTGGGACGGCGGACCGCCGCTTGCCTGG 5803

QY 550 atg-----cccgcttctacgttcgctcccaagatgcttcaagacccgtg 595

Db 5802 AGCTGATCGCGGAAGTCCCGTGTCTGCTTCACAGGCGCGGACGATCTCCAGGACGCT 5743

QY 596 acctgacggcgctcgatggtctctgtggcgcgaggcgctatccattgtccctcgtg 655

Db 5742 TCGTGGCGGCTCGCGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 5683

QY 656 cggaaggtttggtgaggaataatcgccgagcccttggttggggaacttctccgaaaccc 715

Db 5682 CGGAGGGGTTCGCAACCGCGTGGCTCGGGGATGGGCTGGGCGATGGTGTCCCGAGGTCC 5623

QY 716 aagctgctcccatgctaaaagcaggagaagtgtatcctcctcgt---atgagataccattg 772

Db 5622 AGCGGAGCGGTGCTGAGCGAGCGCGCGGTGCTGGCGGCTCGCCCTGAGCCACCGTGG 5563

QY 773 acacaccgtatttggaacgagcgtcgctggaatctagatctctagtagctacacag 832

Db 5562 ACGTCCCGCTGACTGGCAGCAGTGAAGTGAATCGCCCGCGCTGGCGCGGTGCGCGC 5503

QY 833 acgcctgcttgcagcagaatcgagggatt 863

Db 5502 AGCGGTGCGCGGAGCGCGCGGCGGCGACT 5472

RESULT 9

ECRPIAA/c

LOCUS

E.coli genes *iciA* and *rpiA*.

ACCESSION X73026

VERSION X73026.1 GI:405638

KEYWORDS ribose phosphate isomerase A.

SOURCE Escherichia coli.

ORGANISM

BCT

03-NOV-1993

```

Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
REFERENCE
1 (bases 1 to 1804)
AUTHORS Hove-Jensen,B. and Maigaard,M.
TITLE Escherichia coli rpiA gene encoding ribose phosphate isomerase A
JOURNAL J. Bacteriol. 175 (17), 5628-5635 (1993)
MEDLINE 93374859
REFERENCE
2 (bases 1 to 1804)
AUTHORS Hove-Jensen,B.
TITLE Direct Submission
JOURNAL Submitted (23-APR-1993) B. Hove-Jensen, Dept. of Biological
Chemistry, Univ. Inst. of Mol. Biology, Solvgade 83, Dk-1307
Copenhagen, DENMARK
FEATURES
source
1. .1804
/organism="Escherichia coli"
/strain="K12"
/db_xref="taxon:562"
/map="63 min"
/complement(1..529)
/gene="icia"
/complement(<1..529)
/gene="icia"
/codon_start=1
/transl_table=11
/protein_id="CAA51508.1"
/db_xref="GI:405639"
/db_xref="SWISS-PROT:P24194"
/transl_table="MKRDPYRTLOALDAVIRERGFRAAKLCITOSAVSQRKOLEN
MFQQLLVPRVPRPEQCKLLALLRQVELLEEWLQDEQSTPLLLSLVNAADSL
ATWLLPALAPLADSPRLNLQVEDETRTQELRRGEVGVAVSIHQALPSCSLVDRLG
ALDYLFVSSKPFPAEKY"
-10_signal
917..922
/gene="rpiA"
917..1643
/gene="rpiA"
929..1643
/gene="rpiA"
944..946
/gene="rpiA"
957..1616
/gene="rpiA"
/EC_number="5.3.1.6"
/codon_start=1
/transl_table=11
/product="ribose 5-phosphate isomerase"
/protein_id="CAA51509.1"
/db_xref="GI:405640"
/db_xref="SWISS-PROT:P27252"
/transl_table="MTODELKAVGMAALQYVQPTIVGVTGSTAAHFTDALGTMKG
OIEGAVSSSDASTEKLSLGIHVFDLNEVDSLGIYVDGADEINGHMOMIKGGGAALTR
EKTIASVAERFICIDASKQVDILGRFPLPVPVPMARSARQLVKLGRPEYRGV
VTDNGNVILDVHGMELDPLAMENAINAIPGVTVGLFANRGADVALIGTFPDGKTV
K"
terminator
1622..1638
/gene="rpiA"
BASE COUNT 421 a 448 c 477 g 458 t
ORIGIN
Query Match 9.5%; Score 82.6; DB 2; Length 1804;
Best Local Similarity 48.8%; Pred. No. 1.9e-12;
Matches 223; Conservative 0; Mismatches 234; Indels 0; Gaps 0;
Qy 35 tcattgatgaagcagcttcgaagcgccctccttagcccttcattccctcccgcg 94
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 486 TGATACGTGAACGAGGATTGAGCGCGGCGGACAAAAGCTGTGCTACATACAGCCG 427
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 95 tgaatcagcgcttaaaagctctcagcatcacgtgggtcgagtgttgcgcgaccc 154
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 426 TCTCACAGCGCATTAAGCAACTGGAATAATGTTTCGGCGACCGCTGTGTGCGTACCG 367
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 155 aaccgccccaaagcaaccgaagcggtggaagtccctgtgtgaagcgcgcaaatggtgt 214

```

```

| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 366 TACGCCCGCCCGCCGACGCAAGGCAAAAACCTGCTGGCACTGCTGCCAGGTGAGT 307
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 215 tgcgtgaagcagaataaagcgcaactatctggacgccttgctgaaatcccgtaacca 274
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 306 TGCTGGAGAAGAGTGGCTGGCGGATGAACAAACGGTTCGACTCCGCTGTGTTTAC 247
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 275 tcgcatcaacgcagattcgctatccacatggtttccctcccggttcaacagtagctt 334
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 246 TGGCGGTCAACGCCGACAGCTCTGGCGAGCTGTTGCTTCTGCACCTGCTGTTGG 187
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 335 cttggggtgagcaacgctcaacgctgagctggagatgaagcgacacattactctgc 394
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 186 CTGATTGCGCTATCGGCTCAACTTGACAGTAGAAGATGAACCCGCACTCAGGAACGTC 127
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 395 tgcggctggagatgttttagagcggttaaccgcgtgaagctaataccggtggcggaatgt 454
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 126 TGGCGCGCGGGAAGTGGTCGGCGGGTGAGTATTCACATCAGCGCTGCCGAGTTGTC 67
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 455 aagtagtagaacttggaaacctgcgccacttggccat 491
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 66 TTGTCGATAAACTTGGTGCCTCGACTATCTGTTGCT 30
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
RESULT 10
LOCUS EOCOICIA 1815 bp DNA BCT 06-MAR-1995
DEFINITION E.coli chromosome initiation inhibitor (icia) gene, complete cds.
ACCESSION M62865
VERSION M62865.1 GI:146435
KEYWORDS chromosome initiation inhibitor; icia gene.
SOURCE E.coli (strain W3110) DNA, (Kohara library) clone IA2
('miniset'-set No. 471).
ORGANISM Escherichia coli
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
REFERENCE
1 (bases 1 to 1815)
AUTHORS Thony,B., Hwang,D.S., Fradkin,L. and Kornberg,A.
TITLE icia, an Escherichia coli gene encoding a specific inhibitor of
chromosomal initiation of replication in vitro
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 88 (10), 4066-4070 (1991)
MEDLINE 91239508
FEATURES
source
1. .1815
/organism="Escherichia coli"
/sub_strain="W3110"
/strain="K-12"
/db_xref="taxon:562"
/clone="IA2"
/tissue_lib="E.coli Kohara"
/map="62.8 min. on K-12 ma"
650..1543
/gene="icia"
650..1543
/gene="icia"
/codon_start=1
/transl_table=11
/function="Inhibitor of oriC replication in vitro"
/product="chromosome initiation inhibitor"
/protein_id="AA62780.1"
/db_xref="GI:146436"
/transl_table="MKRDPYRTLOALDAVIRERGFRAAKLCITOSAVSQRKOLEN
MFQQLLVPRVPRPEQCKLLALLRQVELLEEWLQDEQSTPLLLSLVNAADSL
ATWLLPALAPLADSPRLNLQVEDETRTQELRRGEVGVAVSIHQALPSCSLVDRLG
ALDYLFVSSKPFPAEKY"
BASE COUNT 468 a 486 c 440 g 421 t
ORIGIN
62.8 min. on K-12 map; kb coordinate 3974.
Query Match 9.5%; Score 82.6; DB 2; Length 1815;
Best Local Similarity 48.8%; Pred. No. 1.9e-12;

```


Db 1684 TGATACGTGAACGAGGATTTGAGCGCGCGCACAAAAGCTGTGATTACAAATCAGCCG 1743
 Qy 95 tgaagcagcggttaaagctctcagcatcacgtgggtcgagtggtgtatcgcgcaccc 154
 Db 1744 TCTCACAGCGCATTAAGCAACTGGAATAATGTTCCGGCAGCCCGCTGTGTGGGTACCG 1803
 Qy 155 aaccggccaaagcaacgaagcgggtgaagtcctgtgcaagcagcgcgaaatgggtg 214
 Db 1804 TACCGCGCGCGCCGACGGAACAGGGCAAAATGCTGGCAGCTGTGCGCAGGTGGAGT 1863
 Qy 215 tgcgtcaagcagaactaaagcgaactatctggaacccctgtgtaaatccggttaacca 274
 Db 1864 TGCTGGAAGAAGAGTGCGTGGCGCATGAACAAACCGTTCGACTCCGCTGCTGCTTTCAC 1923
 Qy 275 tcgcatacaagcagattcgcattcccatggtttccctcccggtttcaacagaggtagctt 334
 Db 1974 TGGCGGTCAACGCCGACAGCTGCGGAGCTGTTGCTTCCTGCAGCTGGCTCTGTGTGG 1983
 Qy 335 cttggggtgagcaacgcctcagtcgcgttggagatgaagcgcacacattactctg 394
 Db 1984 CTGATTGCGCTATCCGCTCACTTGCAGGTAGAAGATGAACCCGCACTCAGGAACGTC 2043
 Qy 395 tgcggtgagatgttttagagcggtaaacccgtgaagctaataccggtgcgagatgtg 454
 Db 2044 TGCGCGCGCGGAAGTGTGCGCGCGGTGAGTATTCACATCAGCGCGCTGCCGAGTTGTC 2103
 Qy 455 aagtagtagaactggaaccatgcccacttggccat 491
 Db 2104 TTGTCGATAACTGGTGGCGCTCGCACTATCTGTTGCT 2140
 RESULT 12
 AE000375
 LOCUS Escherichia coli K12 MG1655 section 265 of 400 of the complete genome. BCT 01-DEC-2000
 DEFINITION
 ACCESSION AE000375 U00096
 VERSION AE000375.1 GI:1789282
 KEYWORDS
 SOURCE Escherichia coli K12.
 ORGANISM Escherichia coli K12.
 Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.
 REFERENCE
 AUTHORS Blattner, F.R., Plunkett, G. III, Bloch, C.A., Perna, N.T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F., Gregor, J., Davis, N.W., Kirkpatrick, H.A., Goeden, M.A., Rose, D.J., Mau, B. and Shao, Y.
 TITLE The complete genome sequence of Escherichia coli K-12
 JOURNAL Science 277 (5331), 1453-1474 (1997)
 MEDLINE 97426617
 PUBMED 9278503
 REFERENCE
 AUTHORS Blattner, F.R.
 TITLE Direct Submission
 JOURNAL Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
 Email: ecoli@genetics.wisc.edu Phone: 608-262-2534 Fax: 608-263-7459
 REFERENCE
 AUTHORS Blattner, F.R.
 TITLE Direct Submission
 JOURNAL Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
 Email: ecoli@genetics.wisc.edu Phone: 608-262-2534 Fax: 608-263-7459
 REFERENCE
 AUTHORS Plunkett, G. III.
 TITLE Direct Submission
 JOURNAL Submitted (13-OCT-1998) Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
 This sequence was determined by the E. coli genome project at the

University of Wisconsin-Madison (Frederick R. Blattner, director). Supported by NIH grants HG0301 and HG01428 (from the Human Genome Project and NCHGR). The entire sequence was independently determined from E. coli K12 strain MG1655. Predicted open reading frames were determined using GeneMark software, kindly supplied by Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA, 30332 [e-mail: mark@amber.gatech.edu]. Open reading frames that have been correlated with genetic loci are being annotated with CG Site Nos., unique ID nos. for the genes in the E. coli Genetic Stock Center (CGSC) database at Yale University, kindly supplied by Mary Berlyn. A public version of the database is accessible (http://cgsc.biology.yale.edu). Annotation of the genome is an ongoing task whose goal is to make the genome sequence more useful by correlating it with other data. Comments to the authors are appreciated. Updated information will be available at the E. coli Genome Project's World Wide Web site (http://www.genetics.wisc.edu). *** The E. coli K12 sequence and its annotations are periodically updated; this is version M54. No sequence changes. Annotation updates: updated gene identifications and products; all new functional assignments courtesy of Monica Riley; added promoters, protein binding sites, and repeated sequences described in reference 1. The unique numeric identifiers beginning with a lowercase 'b' assigned to each gene (protein- or RNA-encoding) are now designated as gene synonyms instead of labels. This should allow them to be searched for in Entrez as gene names.

FEATURES

source
 1. .10362
 /organism="Escherichia coli K12"
 /strain="K12"
 /sub_strain="MG1655"
 /db_xref="taxon:83333"
 complement(9. .40)
 /note="factor Sigma70; predicted +1 start at 3057674"
 34. .63
 /note="factor Sigma70; predicted +1 start at 3057742"
 101. .994
 /gene="icfA"
 /note="b2916"
 101. .994
 /gene="icfA"
 /function="regulator; DNA - replication, repair, restriction/modification"
 /note="b2917; 100 pct identical to ICfA_ECOLI SW: P24194; CG Site No. 28646"
 /codon_start=1
 /transl_table=11
 /product="replication initiation inhibitor, binds to 13-mers at oriC"
 /protein_id="AAC75953.1"
 /db_xref="GI:1789283"
 /translation="MKRPDYRTLQALDAVIRERGFRAAKLCITQSAVSQRIKQLEN MFGQPLIVTPVPPRTEQOKLLALLRQVELLEWGLDQGTSTPLLLSLAVNADSL ATWLLPALAPVLADSPIRNLNLOVEDTRTORRERGVGAVSIQHQALPSCIIVDKLG ALDYLIVSSKPFKAEKYPGCTRSALLKAPVAFDHLDDMHQAFLOQNFDPGCVPC HIYNSSEAFVQLARQGTTCMIPHQLQIEKELASGELLIDLTGFLQRMRLVHWRAPES RMRKVTVDALLDYGKVLROD"
 1045. .1072
 /note="factor Sigma70; predicted +1 start at 3058751"
 1198. .3342
 /gene="sbm"
 /note="b2917"
 1198. .3342
 /gene="sbm"
 /function="enzyme; Degradation of small molecules: Fatty acids"
 /note="o714; 99 pct identical to SBM_ECOLI SW: P27253"
 /codon_start=1
 /transl_table=11
 /product="methylmalonyl-CoA mutase (MCM)"
 /protein_id="AAC75954.1"
 /db_xref="GI:1789284"
 /translation="MSNVQBWQQLANKELSRREKTVDSLIVHQTAEGIAIKPLYTEADL

DNLEVTGTLPLGPPYVGRPRATMYTAQPTWIRQYAGFSTAKESNAYFRNRLAAGOKGL
 SVAEDLATHRGYDSNPRVAGDVGKAGVAIDTVEIMKVLFDQIPIDKMSVSMTMGAV
 LPLAFYIVAAEEQGVTPDKLTQNDILKLEYLCRNTYIYPPKPSMRITADIWAGCS
 GNPFRNTISGYNHGEAGANCVOQVAFTLADIEYIKAAISAGLKIDDFAPRLSFF
 FGIGMDLNNVAMLAARAYLWSEAVSGFGAOPDKSLALRTHOCTQSWISLTEODPNV
 IRTTEALATLGCTQSLHTNAFDEALGLPTDFSARIARTQIIIOESELCTVDPL
 AGSYTESLTDQIVKQARALIOOIDEAGGMAKAIKAGLPKRMIEESAREQLIDOGK
 RVIVGNKYKLDHEDETVDLEIDNVMRNEQIASLERIRATRDAAVTAALNALTHAA
 QHNEILLAAVNAARVATLGEISDALEAFDRYLPVOCVTGVIQAQSHQSEKASE
 FDAIVAOEFOFLADNCRPRILIAKMGODHGRKAVIASAYSDLGFDVLDSPMESTP
 EELARLAVENDVHVGCASSLAAGHKTILPELVEALKKMGREDICVAGGVIPPDYAF
 LQERGVAAIYGCTPMLDSVRDVLNLSOHDD"

gene

/gene="y9fd"

CDS

/note="b2918"

3335..4330

/gene="y9fd"

/function="phenotype; Not classified"

/note="0331; 100 pct identical to yGFD_ECOLI SW:

p27254(257 aa) but has 25 additional N-ter aa 50

additional C-ter aa; orf2 of x66836"

/codon_start=1

/transl_table=11

/product="putative nucleotide-binding protein"

/protein_id="AAC75955.1"

/db_xref="GI:1789285"

/translation="MINATLAESIRRLRQGERATLAQAMTLVESRHRPHQALSTOLL
 DAIMPYNTGLRGVTPGAGKSTFLFAGMLLIREGLKAVIAVDPSPTVGGSL
 GKPTMNDLARAFAIRPVPSSGHLGASORARELMILCEAAGYDVIVETVGVQS
 ETEVARWDCFTSLQIAGGDDLOGIKKMEVADLVINKDDGNHNVAIARHME
 SALHILRQWDFPQVLTCSALERKRGIDEIWHAIIDFKTALTALTAQQVROQOSVE
 WLKQTEEVLDHMFANEDFDYRQTLAVKNNTLSPTRLQLSEFIQOYFD"

gene

/gene="y9fg"

/note="b2919"

4299..5126

/gene="y9fg"

/function="putative enzyme; Not classified"

/note="0275; This 275 aa ORF is 30 pct identical (6 gaps)

to 250 residues of an approx. 296 aa protein ECHM_RAT SW:

p14604"

/codon_start=1

/transl_table=11

/product="putative enzyme"

/protein_id="AAC75956.1"

/db_xref="GI:1789286"

/translation="WNLGRNLLIKIFMSYQYVNVVTINKVAVIEENVGRKLNLSK
 VFIDLMQALDPLNRPEIRCIILRAPSGKVSAGHDHELPSGGRDPLSYDDPLRQI
 TRMIQKFKPILSVWEGSVWGGAFEMINSSDLIIAASSTFSMTFVNLGVYNLVGHI
 NLRDAGEHIVRELIFTASPIAQALAVGLNHNHVEVEELEDFTLQMAHHISEKAPL
 AJAVIKEELRVLGEARTWNSDEFERIQGMRAVYDSYQEGMNAFLKRPNFVGH"
 5150..6628

gene

/gene="y9fh"

/note="b2920"

5150..6628

/gene="y9fh"

/function="putative enzyme; Not classified"

/note="0492; This 492 aa ORF is 46 pct identical (7 gaps)

to 490 residues of an approx. 544 aa protein CATL_CLOKL

SW: p38946"

/codon_start=1

/transl_table=11

/product="putative coenzyme A transferase"

/protein_id="AAC75957.1"

/db_xref="GI:1789287"

/translation="METOWTRMTANAAEIIQHNDMAVFSGFTPAGSPKALPTATARR
 ANEQEAKKPYOIRLLTGTASISAAADDVLSADAVSWRAPYOTSSGLRKKINOGAVF
 VDLHLSVAQWVNYGFGGDIADVIEASALAPDGRVLTSGIGNAPTWLLRAKKVIE
 LNHHDPRVAELADIVIPGAPRNSVSIFHAMDRVGRYVOIDPKKIVAVVETNLPD
 AGNMIDQCMQCOQIADNVVTLFQEMAHGRIPPEFLPQSGVGNINNAVARLENP
 VIPSNNPSEVLSVHLETKISGASASLTSADSLRKIYDNMDYFASRIVLRP
 OEISNNPPIIRRLGVALNGLFEDFYGHANSTHVAGVGLMNGIGSGDFERNVLSI
 FMAPSIAREGKISTVVPWCSHVHDSHSESVKVIITFQGLADLRGLSPLOARTIIDNCA

HPMYRDYLRHYLENAPGGHIHHDLSHVFDLHRNLATGSMGLG"
 complement(6625..7536)
 /gene="y9fi"

/note="b2921"

complement(6625..7536)

/gene="y9fi"

/function="putative regulator; Not classified"

/note="f303; This 303 aa ORF is 25 pct identical (11 gaps)

to 241 residues of an approx. 304 aa protein CYNR_ECOLI

SW: p27111"

/codon_start=1

/transl_table=11

/product="putative transcriptional regulator LYSR-type"

/protein_id="AAC75958.1"

/db_xref="GI:1789288"

Query Match 9.5%; Score 82.6; DB 1; Length 10362;

Best Local Similarity 48.8%; Pred. No. 1.7e-12;

Matches 223; Conservative 0; Mismatches 234; Indels 0; Gaps 0;

QY 35 tcattgatgaagcagcttcgaagcgccctccttagcccttcattcccccctcgcg 94

Db 144 TGATACGTGAACGAGGATTGACGGCGGCGCACAAAAGCTGTCATTACAAATCAGCCG 203

QY 95 tgagtcagcgcgcttaaaagctctcgagcatcaogtgggtcgagtggtgtatcgccacc 154

Db 204 TCTCACAGCGCATTAAGCAACTGGAAAATATGTCGGGACGCGCTGTTGGTCGTACCG 263

QY 155 aaccggcgaagcaaacgaagcggtggaagtccttgcgaagcagcgcggaataatggt 214

Db 264 TACCGCGCGCGCGACGAGGAAACAGGCGCAAAACTGCTGTCACATGCTGCGCAGGTGGAGT 323

QY 215 tgcgtcaagcgaacaaactaaagcgcaactatctgagcgcttgcgaaatcccggttaacca 274

Db 324 TCGTGAAGAAGAGTGGCTGGCGGATGACAAACCGGTTCCGACTCCGCTGCTTTCAC 383

QY 275 tcgcatcaacgcagatctgcctatccacatggttctcccggttgcacagaggtagctt 334

Db 384 TGGCGGTCAACGCCGACAGTCTGGCGACGTGTTGCTGCTGCTGCTGCTGCTGCTGCTG 443

QY 335 cttggggtgagcaagcgtcagctgcgtgctggaagataagcgcaacattatccttc 394

Db 444 CTGATTGCGCTATPCCGCCTCAACTTCAGGTAGAAGATGAACCCGACCTCAGGAACGTC 503

QY 395 tgcggcggtgagatgttttaggagcggttaacccgtgaagcctaaccgtggcggtggtg 454

Db 504 TCGCGCGCGCGGAAGTGGTGGCGCGGTGAGTATTCAACATCAGCGCGTCCGAGTTGTC 563

QY 455 aagtagtagaacttggaacacatgcgccacttgccat 491

Db 564 TTGTCGATAAACTTGTGGCGCTCGACTATCTGTTCTGT 600

RESULT 13

AE005521

LOCUS

DEFINITION

Escherichia coli O157:H7 EDL933 genome, contig 3 of 3, section 140

of 290.

ACCESSION

AE005521

VERSION

AE005521.1

GI:12517435

KEYWORDS

Escherichia coli O157:H7 EDL933.

Escherichia coli O157:H7 EDL933

Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

Escherichia.

REFERENCE

1 (bases 1 to 10367)

Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,

Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,

Postai,G., Hackett,J., Link,S., Boutin,A., Shao,Y., Miller,L.,

Grotbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K.,

Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,

Welch,R.A. and Blattner,F.R.

Genome sequence of enterohaemorrhagic Escherichia coli O157:H7

JOURNAL Nature 409 (6819), 529-533 (2001)
MEDLINE 21074935
PUBMED 11206551
REFERENCE 2 (bases 1 to 10367)
AUTHORS Perna, N.T., Plunkett, G. III, Burland, V., Mau, B., Glasner, J.D., Rose, D.J., Mayhew, G.F., Evans, P.S., Gregor, J., Kirkpatrick, H.A., Posfai, G., Hackett, J., Klink, S., Boutin, A., Shao, Y., Miller, L., Grotbeck, E.J., Davis, N.W., Lim, A., Dimalanta, E., Potamousis, K., Apodaca, J., Anantharaman, T.S., Lin, J., Yen, G., Schwartz, D.C., Welch, R.A. and Blattner, F.R.

TITLE Direct Submission
JOURNAL Submitted (22-OCT-2000) Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
FEATURES Location/Qualifiers
source 1..10367
/organism="Escherichia coli O157:H7 EDL933"
/strain="EDL933"
/serotype="O157:H7"
/db_xref="taxon:155864"
/note="enterohemorrhagic"
complement(357..1559)
/genes="visc"
/note="24243"
complement(357..1559)
/genes="visc"
/function="orf; Unknown function"
/note="Residues 1 to 400 of 400 are 98.75 pct identical to residues 1 to 400 of 400 from Escherichia coli K-12 Strain MG1655; B2906"
/codon_start=1
/transl_table=11
/product="orf, hypothetical protein"
/protein_id="AAG58033.1"
/db_xref="GI:12517436"
/translation="MOSVDVAIVGGMGVLAVAGLGGSLRVAVLEORVPEPLAADA PPOLRYSAINAAEKLILRLGVODILSRASCYHGMEDKDSFGHIFSDQSMGYS HLGHIVENSIVHIALNKAQOOSDITLLAPAELOQVANGENETFLIKGSGMTARLV IGADGANSJURNKADILPTFDYOHHALVATIRTEPHDAVARQVFHGEIGLAFPLPS DPHLCISVWSLSPPEAQMOQASEDFNRLNATFONRLGCRKVESARQVFPITGRYA ROPAAHRLALVGAHTIHPLAGQGVNLGFMADAEILAEKRLHROGKDGIOYILYLR YRSRKHSAALMAGMQGFRDLFSGTNPVKKLLRDIGLKLADTLPGVKQLIRQAMGL NDLPENLR"
complement(1583..2761)
/genes="ubiH"
/note="24244"
complement(1583..2761)
/genes="ubiH"
/function="enzyme; Biosynthesis of cofactors, carriers: Menaquinone, ubiquinone"
/note="Residues 1 to 392 of 392 are 98.97 pct identical to residues 1 to 392 of 392 from Escherichia coli K-12 Strain MG1655; B2907"
/codon_start=1
/transl_table=11
/product="2-octaprenyl-6-methoxyphenol--> 2-octaprenyl-6-methoxy-1, 4-benzoquinone"
/protein_id="AAG58034.1"
/db_xref="GI:12517437"
/translation="MSVLIIVGGMGAGATLALISRLSHGALPVHLIEATAPESHHPG FGRALTAGTCQQLARIGWQSLADCATITTVSDRGHAGFVTLAEADYQLAAL GOVELHNVGQRFALLRKAPGYTLHCPDRVANVARTQSHVEVTLEGGETLGRVLVA RCTGSHALATVCGVDWQOEPYEQLAIVANVATVAHGRAFRFTQHGPLMLPMSDG RCLVWCHPLEREEVLSWDEKFCRELQSAFGWRLGKITHAGKRSAYPLALTHAARS ITHRTVLVGNAAQTLHPIAGOGFNLGMRDVMSLAETLTOAQERGEDMDGYGLCRYQQ RQSDREATIGVTDLSLVHLFANRWTFIVVGRNIGLTMELFTPARDVLAQRYXGWYAR "
complement(2758..4083)
/genes="pepp"
/note="24245"
complement(2758..4083)
/genes="pepp"
/function="enzyme; Macromolecule degradation: Degradation of proteins, peptides, glyco"

gene
/note="Residues 1 to 441 of 441 are 99.09 pct identical to residues 1 to 441 of 441 from Escherichia coli K-12 Strain MG1655; B2908"
/codon_start=1
/transl_table=11
/product="proline aminopeptidase P II"
/protein_id="AAG58035.1"
/db_xref="GI:12517438"
/translation="MSEISRQEFQRRRQALVEQMFGSAALIFAAPVTRSDSEYPY RQNSDFWYTFGFNEPEAVLVLIKSDDTHHNSVLFNRVRLTAEIFWGRRLGQDAPEK LGVDALAFSEINQQLYQLLGLDVVYHAQGEYAYADEIVNSALEKRLKSGRONLTAP ATMDWRPVVHEMRLEKSPETIAVLRAGEITAMATRAMKRCPCMEVHLEGETIHH EFNRHGARTSYNTIVSGGNGCILHYTENECEMRDGLVIDAGCEYKGYAGDIIRT FPNVKGFTQAQREIYDIVLESUETSLRLYRPGYSIILEVTEGVVRIMVSGVLKGLIKG DIDELIAQNAHRFFPMHGLSHWGLDHDVGVYQDRSRILEPGMVLTVPEPGLYIAPD ADVPEQYRGIGRIEDDIIVITETNGENLTASVVKKPEEIEALMAAARKQ"
complement(4109..4693)
/genes="ygfB"
/note="24246"
complement(4109..4693)
/genes="ygfB"
/function="orf; Unknown function"
/note="Residues 1 to 194 of 194 are 100.00 pct identical to residues 1 to 194 of 194 from Escherichia coli K-12 Strain MG1655; B2909"
/codon_start=1
/transl_table=11
/product="orf, hypothetical protein"
/protein_id="AAG58036.1"
/db_xref="GI:12517439"
/translation="MLMSIQENMPGYNEMNOYLNQOQGLTFAEMHGLISGMICGGND DSSWPLLLHDLTNEGMAFGHLEAQAQAKRMHSATSDALQDDGFLFQLYLPDGDVDFD RADLAGWVNHFLGLGVTPQKLDKVTGETGEADIDRLNIAQLGYDEDEDEELEMSL EETIEIVRVAALICHDTFTHPQTAPEVQKPTLH"
4855..5184
/genes="ygfE"
/note="24247"
4855..5184
/genes="ygfE"
/function="orf; Unknown function"
/note="Residues 1 to 109 of 109 are 100.00 pct identical to residues 1 to 109 of 109 from Escherichia coli K-12 Strain MG1655; B2910"
/codon_start=1
/transl_table=11
/product="orf, hypothetical protein"
/protein_id="AAG58037.1"
/db_xref="GI:12517440"
/translation="MSAQPDVQIFGRSLRVNCPDPORDALNQAADLNRLODLKER TRVTNTEQVLVFTAAALNISYELAQEKARTRDYAASMEQIRMIQQTIEQALLBQGRITE KTNQNP"
5225..5408
/genes="ssrS"
/product="6S RNA"
/function="RNA; Macromolecule synthesis, modification: Ribosomal and stable RNAs"
5225..5408
/genes="ssrS"
/note="24248"
5484..6032
/genes="ygfA"
/note="24249"
5484..6032
/genes="ygfA"
/function="putative enzyme; Not classified"
/note="Residues 1 to 182 of 182 are 99.45 pct identical to residues 1 to 182 of 182 from Escherichia coli K-12 Strain MG1655; B2912"
/codon_start=1
/transl_table=11
/product="putative ligase"
/protein_id="AAG58038.1"
/db_xref="GI:12517441"

misc_rna
/product="6S RNA"
/function="RNA; Macromolecule synthesis, modification: Ribosomal and stable RNAs"
5225..5408
/genes="ssrS"
/note="24248"
5484..6032
/genes="ygfA"
/note="24249"
5484..6032
/genes="ygfA"
/function="putative enzyme; Not classified"
/note="Residues 1 to 182 of 182 are 99.45 pct identical to residues 1 to 182 of 182 from Escherichia coli K-12 Strain MG1655; B2912"
/codon_start=1
/transl_table=11
/product="putative ligase"
/protein_id="AAG58038.1"
/db_xref="GI:12517441"


```
QY 215 tgcgtcaagcagaactaaagcgcaactatctgacgccttctgacgcttgcgaaatcccggttaacca 274
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 96729 TGCTGAAGAAGAGTGGCTGGCGGATGAAACAAACCGGTTTCGACTCGCTGCTGCTTTCAC 96670
QY 275 tgcgcataacgcagattcgctatcacatggtttctcccggtttcaacgaggtagctt 334
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 96669 TGGCGGTCAACGCCGACGAGCTGGCGACGTTGGTCTCTGCTGCTGCTGCTGCTGCTGCTG 96610
QY 335 ctgggggtggagcaagcgtcacgcgtcgcttggaagatgaagcgacacattatctctgc 394
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 96609 CTGATTGCGCTATCCGCTCACTTCAGGTAGATGAATCAACCCGCACTCAGGAACGTC 96550
QY 395 tgcggcgtggagatttttaggagcgtgaacccgtgaagctaatcccggtggcggtgagtg 454
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 96549 TGGCGCGCGCGGAAGTGGTGGCGCGGTGAGTATTCAACATCAGGCGCTCCGAGTTGTC 96490
QY 455 aagtagtaacttggaacatgcgcatttggccat 491
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 96489 TTGTCGATAAACTTGGTGGCGCTCGACTATCTGTTGCT 96453

RESULT 15
LOCUS ECU28377 141744 bp DNA BCT 01-JUL-1995
DEFINITION Escherichia coli K-12 genome; approximately 65 to 68 minutes.
ACCESSION U28377
VERSION U28377.1 GI:882431
KEYWORDS
SOURCE Escherichia coli.
ORGANISM Escherichia coli.
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
1 (bases 1 to 87107)
Plunkett.G.
Direct Submission
Submitted (01-JUN-1995) Guy Plunkett, Laboratory of Genetics,
University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
This sequence was determined as part of the E. coli Genome Project
(Frederick R. Blattner, director) at the University of
Wisconsin-Madison. Supported by award HG00301 from the NIH Human
Genome Project. The entire sequence was independently determined
from E. coli MG1655; overlaps with other sequence determinations
are annotated. This entry should be considered somewhat
provisional; it will be updated and merged with others at a later
date.
FEATURES
Location/Qualifiers
1..141744
/organism="Escherichia coli"
/strain="K-12"
/sub_strain="MG1655"
/db_xref="taxon:562"
/maps="approximately 65 to 68 minutes"
/note="This sequence comprises the following lambda
clones: DD656(EC18-164), DD658(EC17-126), DD660(EC30MM49),
DD663(EC19-62), DD665(EC27-689d), DD667(EC27-357),
DD670(EC30-45), DD674(EC27-1001), DD676(EC27-1160),
DD678(EC30K505-2), DD680(EC30K506-2), and DD682(EC17-140);
M13Janus vector was used for subcloning"
CDS
complement(1..292)
/note="ORF_f97p"
/codon_start=1
/transl_table=11
/protein_id="AAA69070.1"
/db_xref="GI:882432"
/translation="MAIALVTGSGRGIGRATALLAQBYIVAVNYQNLRHAQEVNM
LITQAGGKAFVLAQIDENQVAMFTAIQDHEPLAALVNNAGILFTQCTVE"
complement(1..5335)
/note="similar to GenBank Accession Number X73958
(ECGCVOB)"
misc_feature
complement(70..324)
/note="similar to GenBank Accession Number D21144
(ECOPYU18)"
complement(318..3634)
/note="similar to GenBank Accession Number L20872
(ECOCVHP); bases 3318 to 3438 differ from the sequence
presented here, as well as from X73958 and D21144"
complement(559..3432)
/gene="gcvp"
complement(559..3432)
/gene="gcvp"
/note="ORF_f957"
/codon_start=1
/transl_table=11
/protein_id="AAA69071.1"
/db_xref="GI:882433"
/translation="MTQTLSQLNSGAFIERHIGDPAQAQQOQMLNAVQAQSLNALTGO
IVPKDIQIATPQVGAPEATYEAALAEKAIASRNKRFTSYIGMGTAVQLPVLIRNM
LENQWYATYTPQPEVSGQRLALLNFQVTLDTGLDMSASILLDRATAAEAMAM
AKRYXKLNANREFVASDVHPQTLVDVTRATETFGFEVIVDDAKVLDHODVFGVLLQ
QVGTGEIHDYALISELSKRIIVVSAADIMALLVLLTAPGKGADIVFGSAGFQV
MGYGGPAAFFAAKDEYKRSMFGRIGVSKDAAGNTALRMAQOTREQHIRREKANSI
CTSQVLLANIASLYAVYHGPVGLKRIANRIHRLTDLAAGLQKGLKRLHARHYFTLC
VEVDKAGVLTAAEAELNRSIDLNAVGITLDETTTRENVMHSLFNVLGDNHGLDID
TLDDKVAHDSRSIOAPMLRDDIELTHPVFNRYHSETEMRYMHSILDLALNOAMIP
LGCTMKLNAAEMIPITWPEAELHPFCPEQAEGYQOQMIQAOLADWLKLTGYDAVC
MOPNSGAQGEYAGLAIARHIESRNEGRDICIIPASAGTNPASAHGAMGQVYVAC
DKNGNIDLTDLRAKAEQADNLSCIMVTYPTSTHYEETIREVCEVHQFGQVYLDG
ANMAQVGTISPGFIDGVSHLNHKFTCIPIHGGGPGMGPIGVKHAFLAPFVPGHVV
QIEGMLTQGVASAAFGSAILPIPSMWYIRMMGAEGKSKAQSAKAILNANYIASRLQD
APVLYTGRDGRVAHECILDTRPLEETGISELDIAKRLIDYGHAPTMSRPVACTLM
VEPTESKVELDREIDAMLAIRAEIDQVKGAVPLEDNPLVNAFHIOSELVAEWAHP
YSREVAFPAGVADKYWTVKRLDDVYGRNLFCSQVPISEYQ"
complement(3551..3940)
/gene="gcvh"
complement(3551..3940)
/gene="gcvh"
/note="ORF_f129"
/codon_start=1
/transl_table=11
/protein_id="AAA69072.1"
/db_xref="GI:882434"
/translation="MSNVPAELKYSKEHLRKEADGTYTVGITEHAQELLDGMVFVD
LPEVCATVSAGDCAVAFSVAASDIYAPVSGEIVAVNDALSDSPELYNSEPYAGGWI
FKIKASDSELESLLDAYEALLEDE"
complement(3809..5631)
/note="similar to GenBank Accession Number M97263
(ECOTPRO)"
complement(3964..5058)
/gene="gcvh"
complement(3964..5058)
/gene="gcvh"
/note="ORF_f364"
/codon_start=1
/transl_table=11
/protein_id="AAA69073.1"
/db_xref="GI:882435"
/translation="MAOQTPLYEQHTLCGARWDPFGWMMPLHYQSQIDEHAVRTDA
GMFVDSHTIYDLRGSRTRFLRYLLANDVAKLTRKSGKALYSGMLNAGSGVDDLIYV
YTFEFLFVNSATREKDLGWITQHAEPFGFIEIVRDDLSIAVQGNQAQKATLF
NDAQOAVGEMKMPFGVQAGDLFIATTGTGTAEGYIALPNEKAADFWRALVEAGVKP
CGLGARDTLRLKAGNLYQSEMDTISPAAANMGWTIAWEPADRFIGREALEVQREH
GTEKLVLGVMTEKGVRLNELPVRFVTRNRGKAVA"
complement(5370..10317)
/note="similar to GenBank Accession Number D90281
(ECOPUV)"
complement(5506..6708)
/gene="visc"
complement(5506..6708)
/gene="visc"
/note="ORF_f400"
/codon_start=1
/transl_table=11
/protein_id="AAA69074.1"
/db_xref="GI:882436"
misc_feature
```

/translation="MQSDVVAIVGGVHGLAVACGLOGSLRVAVLQRVQVPEPLAANA
PQURVSAINAESKLRLGLGVDODIILSRASCYHGMVMDKDSFGHISFDDQSMGY5
HLGHVSIWIHYALMKAHQSSDITLLAPAELOQVAMGENETELTKDGSMLTARLV
IGDAGANSWLNRKADIDPLTWDYQHALVATIRTEPHDAVARQVHGEGLTAPLPS
DPLCSLWSPPEAQMOQASDEFENRALNIAFDNRKLCKVESARQVFPPLTRYA
ROFASHRLALVGDAHTIPLAGOGVNLGFMDAAELIAELKRLHROCKDIGOYLYLR
YERSKRISALMLAGMOGFRDLFSGTNPAPKLLRDIGLKLADTLPGVKPQLIROMGL
NDLPFWLR"
complement(6731..7909)
/gene="ubiH"
complement(6731..7909)
/gene="ubiH"
/note="ORF_f392"
/codon_start=1
/transl_table=11
/db_xref="GI:882437"
/protein_id="AAA69075.1"
/translation="MSVLIYGGVAGATLALISRLSHGALPVHLIEATPESHAPG
FDGRAIALAGTCQQLARIGVWQSLACATAITIVVSDRHAGFVTLAEDYDQLAAL
GQVVELHNGORLFAALLKAPGVTLHCPDRVANVARQSHVEVTELTGRVLVA
ADGTHSALATACGYDQOEPEYEQLAIVANVATSAHEGRAFERFTQHGPLAMLPMSDG
RCSLVWCHPLERREVEVLSWDEKFCRELQSAFGWRLGKIITHAGKRSPYPLALTHAARS
ITHRTVLVGNAAOTLHPFIAOGGNLGMHMDVMSLAETLTOAQERGEDMDGYGLCRYQO
RROSDREATICVTDLSLHVFANRWAPLVVGNIGLMTMELTTPARDVLAQRTLGWVAR
"
complement(7749..9599)
/note="similar to GenBank Accession Number D00398
(ECOAPP2)"
complement(7906..9231)
/gene="pepp"
complement(7906..9231)
/gene="pepp"
/note="ORF_f441; third start codon"
/codon_start=1
/transl_table=11
/protein_id="AAA69076.1"
/db_xref="GI:882438"
/translation="MSLSIQEQRORRRQALVEQMOPGSAALIFAAPVTRSADEYFY
RQNSDFWYTFNFEPEAVLVLIKSDTHNHSVLFRNVRDLTAEIFRFRRLQGDAAPEK
LGYDRALFSEIQOLLYOLLGVVYHAGEXAYADVIYNSALEKLRKGSQNLITAP
ATMDWRPVVHEMLFKSPFEIAYLRRAGETAMAHTRAMEKCRPGCFHLEGEIHH
EFNRHGARYPSYNTIVGSGENGCIILHYTENECMRQGDVLIIDAGCEYKYAGDITRT
FPVNGKFTQAKRELYDIVLESLETSRLYRPGTSILEYTGVEVVRIMVSLVKLILKG
DVBLIAQNAHRFFMHGLSHWGLDHDVYTGQDRSRILEPGMVLVTEPGLTIAPD
AEVPEQYRGIGRIEDDITVITGNGENLTASVVKPPEEIALMVAARKQ"
complement(9257..9841)
/gene="ygfB"
complement(9257..9841)
/gene="ygfB"
/note="ORF_f194"
/codon_start=1
/transl_table=11
/protein_id="AAA69077.1"
/db_xref="GI:882439"
/translation="MLMSIONEMPGYNEMNOYLNQOGTGLTPAEMHGLISGMICGGND
DSSWPLLLHDLTNEGHAFHGLAQALRKMHSATSDALQDDGFLFLYLPQDDVSVFD
RADALAGVNHFLIGLGVTPQKLDKQVTGETGEAIDDUURNIAQLGYDEDEDEQEELEMSL
EETIEYVRVAALLCHDTFTHPQTAPEVQKPTLH"
10003..10332
/note="ORF_o109"
/codon_start=1
/transl_table=11
/protein_id="AAA69078.1"
/db_xref="GI:882440"
/translation="MSAQPVYDIQIFGRSLRVNCPDQORDALNQAAADLNORLODKER
TRVNTQQLVFIAAALNISYELAQEAKTRDYAASMEQRIQLQOTIEQALLEQGRITE
KTQNFPE"
10266..11219
/note="similar to GenBank Accession Number M12965
(ECOSSR)"
10374..10556
/gene="ssr"
/product="stable 6S RNA"

gene 10374..10556
/gene="ssr"
10374..10557
/note="similar to GenBank Accession Number X01238
(ECRRNG6)"
Query Match 9.5%; Score 82.6; DB 2; Length 141744;
Best Local Similarity 48.8%; Pred. No. 1.4e-12;
Matches 223; Conservative 0; Mismatches 234; Indels 0; Gaps 0;
Qy 35 tcattgtaaggcagcttcgaagcgctcttagccttccattcccccctcgcgcg 94
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 14187 TGATACGTGACAGGAGATTGAGCGCGGCGCACAAAACCTGTGCATTACAAATCAGCGG 14246
Qy 95 tgaagtcagcgcttaagcctctcgagcatcacgtgggtcgaggtgttggtatcgcgaccc 154
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 14247 TCTCACAGCGCATTAAGCAACTGGAAATAATGTTCCGGCAGCGCGTGTGGTGCGTACCG 14306
Qy 155 aaccggccaaagcaccgaacgaagcggtgaaagtccttgatgcaagcagcgcggaataatggtgt 214
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 14307 TACCGCCGCGCCGACGGAACAAGGGCAAAAACCTGCTGGCACGTGCTGCCGCCAGGTGGAGT 14366
Qy 215 tgctgcaagcagaaactaaagcgcaactatctggagcctcttgctgaaatcccgcttaacca 274
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 14367 TGCTGGAAGAAGAGTGGCTGGCGGATGAACAAACCGGTTCCACTCCGCTGCTGCTTTCAC 14426
Qy 275 tcgcatcaacgcagattcgctatccacatggtttctctcccggtgttcacacgaggtagctt 334
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 14427 TGGCGGTCAACGCCGCACAGTCTGGCGAGCTGGTGTCTCTCGACTGGCTGCTGTGTTGG 14486
Qy 335 cttgggggtggagcaacgcctcagctgcgcttggaagatgaagcagcattatcccttgc 394
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 14487 CTGATTCCGCTATCCGCGCTCAACTTGCAGGTAGAAGATGAACCCGCACTCAGGAACGTC 14546
Qy 395 tgcggcggtggagatgttttagagcggttaaccccggtgaagctaatcccggtggcggtatgtg 454
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 14547 TGGCGCGCGGGAAGTGGTGGCGCGGTGAGTATTCAACATCAGGCGCTGCCGAGTTGTC 14606
Qy 455 aagtagtagaacttggaaacctgcccacttggccat 491
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 14607 TTGTCGATAAACTTGGTGGCGCTCGACTATCTGTTTCGT 14643

Search completed: September 14, 2001, 07:55:07
Job time: 15209 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 14, 2001, 07:54:24 ; Search time 171.99 Seconds
(without alignments)
3187.151 Million cell updates/sec

Title: us-09-105-117i-1-copy_1421_2293
Perfect score: 873
Sequence: 1 atgacccattcaactgga.....tcgaggattgcggccttag 873

Scoring table:
IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 730101 seqs, 313950809 residues

Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_0601:*

1:	/cgnl_9/gcgdata/geneseq/geneseq/NA1980.DAT:*
2:	/cgnl_9/gcgdata/geneseq/geneseq/NA1981.DAT:*
3:	/cgnl_9/gcgdata/geneseq/geneseq/NA1982.DAT:*
4:	/cgnl_9/gcgdata/geneseq/geneseq/NA1983.DAT:*
5:	/cgnl_9/gcgdata/geneseq/geneseq/NA1984.DAT:*
6:	/cgnl_9/gcgdata/geneseq/geneseq/NA1985.DAT:*
7:	/cgnl_9/gcgdata/geneseq/geneseq/NA1986.DAT:*
8:	/cgnl_9/gcgdata/geneseq/geneseq/NA1987.DAT:*
9:	/cgnl_9/gcgdata/geneseq/geneseq/NA1988.DAT:*
10:	/cgnl_9/gcgdata/geneseq/geneseq/NA1989.DAT:*
11:	/cgnl_9/gcgdata/geneseq/geneseq/NA1990.DAT:*
12:	/cgnl_9/gcgdata/geneseq/geneseq/NA1991.DAT:*
13:	/cgnl_9/gcgdata/geneseq/geneseq/NA1992.DAT:*
14:	/cgnl_9/gcgdata/geneseq/geneseq/NA1993.DAT:*
15:	/cgnl_9/gcgdata/geneseq/geneseq/NA1994.DAT:*
16:	/cgnl_9/gcgdata/geneseq/geneseq/NA1995.DAT:*
17:	/cgnl_9/gcgdata/geneseq/geneseq/NA1996.DAT:*
18:	/cgnl_9/gcgdata/geneseq/geneseq/NA1997.DAT:*
19:	/cgnl_9/gcgdata/geneseq/geneseq/NA1998.DAT:*
20:	/cgnl_9/gcgdata/geneseq/geneseq/NA1999.DAT:*
21:	/cgnl_9/gcgdata/geneseq/geneseq/NA2000.DAT:*
22:	/cgnl_9/gcgdata/geneseq/geneseq/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	873	100.0	993	22 AAF71777	Corynebacterium gl
2	873	100.0	2374	18 AAT96816	DNA encoding lysG,
3	122.6	14.0	15239	17 AAT33536	BCG deletion regio
4	45.4	5.2	45613	22 AAF28535	Genomic fragment #
5	45.2	5.2	4915	20 AAX84317	Stealth virus nucl
6	41.6	4.8	2842	21 AAD00334	Rice raffinose syn
7	37	4.2	2058	9 AAN80093	Sequence of BamHI/
8	37	4.2	2058	16 AAQ97857	P. aeruginosa tfda
9	36.8	4.2	102634	21 AAAB1464	N. meningitidis pa
10	36.8	4.2	349980	21 AAF21609	Neisseria meningit
11	35	4.0	4689	21 AAZ87299	S. venezuelae macr

ALIGNMENTS

RESULT 1

AAAF71777					
ID	AAAF71777	standard; DNA; 993 BP.			
XX	AAAF71777;				
AC	AAAF71777;				
DT	30-APR-2001	(first entry)			
DE	Corynebacterium glutamicum MP protein nucleotide sequence SEQ ID NO:49.				
XX					
XX	Corynebacterium glutamicum; metabolic pathway protein; MP protein;				
XX	fine chemical production; microorganism; organic acid; nucleoside;				
XX	nonproteinogenic amino acid; purine base; pyrimidine base; nucleotide;				
XX	lipid; saturated fatty acid; unsaturated fatty acid; diol; vitamin;				
XX	carbohydrate; aromatic compound; cofactor; polyketide; enzyme; ds.				
OS	Corynebacterium glutamicum.				
XX					
XX	WO200100843-A2.				
PN					
PD	04-JAN-2001.				
XX					
PF	23-JUN-2000; 2000WO-IB00923.				
XX					
XX	25-JUN-1999; 99US-0141031.				
PR	01-JUL-1999; 99DE-1030476.				
PR	02-JUL-1999; 99US-0142101.				
PR	08-JUL-1999; 99DE-1031415.				
PR	08-JUL-1999; 99DE-1031418.				
PR	08-JUL-1999; 99DE-1031419.				
PR	08-JUL-1999; 99DE-1031420.				
PR	08-JUL-1999; 99DE-1031424.				
PR	08-JUL-1999; 99DE-1031428.				
PR	08-JUL-1999; 99DE-1031434.				

S. venezuelae pik
S. venezuelae pik
Nucleotide sequenc
Recombinant cosmid
Sequence encoding
Corynebacterium gl
Corynebacterium gl
Human TGF-beta act
Human TGF-beta act
Human TGF-beta act
Polynucleotide seq
Human B7-1 exon 5.
B7 antigen gene.
Human B lymphocyte
Human B7 CDNA sequ
Human B7-2 antigen
Human B lymphocyte
Human B7 CDNA. HO
Human B7.1 CDNA.
Human B7.1 coding
Human B lymphocyte
Human TAK1-6xHis e
Human transforming
Human TAK1 encodin
Human TAK-1 nucleo
Human calcium chan
Human mucosal adre
Human mucosal adre
Human mucosal adre
cDNA encoding huma
Human low adenosin
Human adenosine re
Corn raffinose syn
Human low adenosin

Res 1 (AAT-7R2)

(LypG)

PR 08-JUL-1999; 99DE-1031435.
 PR 08-JUL-1999; 99DE-1031443.
 PR 08-JUL-1999; 99DE-1031453.
 PR 08-JUL-1999; 99DE-1031457.
 PR 08-JUL-1999; 99DE-1031465.
 PR 08-JUL-1999; 99DE-1031478.
 PR 08-JUL-1999; 99DE-1031510.
 PR 08-JUL-1999; 99DE-1031541.
 PR 08-JUL-1999; 99DE-1031573.
 PR 08-JUL-1999; 99DE-1031592.
 PR 08-JUL-1999; 99DE-1031632.
 PR 08-JUL-1999; 99DE-1031634.
 PR 08-JUL-1999; 99DE-1031636.
 PR 09-JUL-1999; 99DE-1032125.
 PR 09-JUL-1999; 99DE-1032126.
 PR 09-JUL-1999; 99DE-1032130.
 PR 09-JUL-1999; 99DE-1032186.
 PR 09-JUL-1999; 99DE-1032206.
 PR 09-JUL-1999; 99DE-1032227.
 PR 09-JUL-1999; 99DE-1032228.
 PR 09-JUL-1999; 99DE-1032229.
 PR 09-JUL-1999; 99DE-1032230.
 PR 14-JUL-1999; 99DE-1032922.
 PR 14-JUL-1999; 99DE-1032926.
 PR 14-JUL-1999; 99DE-1032928.
 PR 14-JUL-1999; 99DE-1033004.
 PR 14-JUL-1999; 99DE-1033005.
 PR 14-JUL-1999; 99DE-1033006.
 PR 12-AUG-1999; 99US-0148613.
 PR 27-AUG-1999; 99DE-1040764.
 PR 27-AUG-1999; 99DE-1040765.
 PR 27-AUG-1999; 99DE-1040766.
 PR 27-AUG-1999; 99DE-1040832.
 PR 31-AUG-1999; 99DE-1041378.
 PR 31-AUG-1999; 99DE-1041379.
 PR 31-AUG-1999; 99DE-1041380.
 PR 31-AUG-1999; 99DE-1041394.
 PR 31-AUG-1999; 99DE-1041396.
 PR 03-SEP-1999; 99DE-1042076.
 PR 03-SEP-1999; 99DE-1042077.
 PR 03-SEP-1999; 99DE-1042079.
 PR 03-SEP-1999; 99DE-1042086.
 PR 03-SEP-1999; 99DE-1042087.
 PR 03-SEP-1999; 99DE-1042088.
 PR 03-SEP-1999; 99DE-1042095.
 PR 03-SEP-1999; 99DE-1042124.
 PR 03-SEP-1999; 99DE-1042129.
 PR 09-MAR-2000; 2000US-0187970.
 PA (BADI) BASF AG.

XX Pompejus M, Kroeger B, Schroeder H, Zelder O, Haberhauer G;
 PI WPI: 2001-137957/14.
 DR P-PSDB; AAB79658.
 XX
 PT Nucleic acids from *Corynebacterium glutamicum* encoding metabolic
 PT pathway proteins, useful for producing fine chemicals in
 PT microorganisms, including organic acids, nonproteinogenic amino acids,
 PT and purine and pyrimidine bases -
 XX

XX Claim 3; Page 226-228; 1737pp; English.

XX AAF1753 to AAF72330 encode the *Corynebacterium glutamicum* metabolic
 CC pathway (MP) proteins given in AAB79634 to AAB80211. The C. glutamicum
 CC MP nucleic acids are useful for the production of fine chemicals
 CC in microorganisms, including organic acids, nonproteinogenic amino
 CC acids, purine and pyrimidine bases, nucleosides, nucleotides, lipids,
 CC saturated and unsaturated fatty acids, diols, carbohydrates, aromatic
 CC compounds, vitamins, cofactors, polyketides and enzymes.
 XX

SQ Sequence 993 BP; 233 A; 256 C; 268 G; 236 T; 0 other;

		Query Match	100.0%;	Score 873;	DB 22;	Length 993;
		Best Local Similarity	100.0%;	Pred. No. 1.9e-268;		
Matches	873;	Conservative	0;	Mismatches	0;	Indels
Gaps	0;					
QY	1	atgaacccattcaactgacactttgctctcaatcattgatgaagcgagcttcgaagc	60			
DB	101	atgaacccattcaactgacactttgctctcaatcattgatgaagcgagcttcgaagc	160			
QY	61	gcctcttagcctttccattccctcgcggtgagtcagcgcttaaaagctctcgag	120			
DB	161	gcctcttagcctttccattccctcgcggtgagtcagcgcttaaaagctctcgag	220			
QY	121	catcacgtgggtcgagtggttgcgcaccccaaccgccaagaacccgaagcggt	180			
DB	221	catcacgtgggtcgagtggttgcgcaccccaaccgccaagaacccgaagcggt	280			
QY	181	gaagtccttgcgaagcagcgcggaatgtgttctcgaagcagaactaaagcgaa	240			
DB	281	gaagtccttgcgaagcagcgcggaatgtgttctcgaagcagaactaaagcgaa	340			
QY	241	ctatctggagccttgcgaatcccggttaaccatcgccatcaacgagattcgctacc	300			
DB	341	ctatctggagccttgcgaatcccggttaaccatcgccatcaacgagattcgctacc	400			
QY	301	acatggtttctccctcgtgttcaacgaggtagcttcttgggtgagcaacgcgtcacgtg	360			
DB	401	acatggtttctccctcgtgttcaacgaggtagcttcttgggtgagcaacgcgtcacgtg	460			
QY	361	cgcttggaagatgaagcgaacacattatccttgcgtcggcggtgagatgttttagagcg	420			
DB	461	cgcttggaagatgaagcgaacacattatccttgcgtcggcggtgagatgttttagagcg	520			
QY	421	gtaacccgtgaagtaataccgtgcggtgagatgtgaagtagaacttgaacacgcgc	480			
DB	521	gtaacccgtgaagtaataccgtgcggtgagatgtgaagtagaacttgaacacgcgc	580			
QY	481	cacttgccattgcaacccctcctcattgcgggagtcctacatggttgatgggaactagat	540			
DB	581	cacttgccattgcaacccctcctcattgcgggagtcctacatggttgatgggaactagat	640			
QY	541	tgggtgcgactcccgctttagcgttcggtcccaagatgtctcaagacgtgacctg	600			
DB	641	tgggtgcgactcccgctttagcgttcggtcccaagatgtctcaagacgtgacctg	700			
QY	601	gacggcgcgctcgatgctctgtggggcgagcggtatccatttcccgctcgcgga	660			
DB	701	gacggcgcgctcgatgctctgtggggcgagcggtatccatttcccgctcgcgga	760			
QY	661	ggttttggtgaggaattcgcgcgagccttgggtgggagcttctcccgaaaccccaagct	720			
DB	761	ggttttggtgaggaattcgcgcgagccttgggtgggagcttctcccgaaaccccaagct	820			
QY	721	gctccatgctaaaagcaggaagatgctcctcctcgatgagatccattacacacg	780			
DB	821	gctccatgctaaaagcaggaagatgctcctcctcgatgagatccattacacacg	880			
QY	781	atgtattgcaacgatggcgcttgggaatctcttagctcttagctcacagcgctc	840			
DB	881	atgtattgcaacgatggcgcttgggaatctcttagctcttagctcacagcgctc	940			
QY	841	gttgatgcagcaatcgagggtggtggtggttag	873			
DB	941	gttgatgcagcaatcgagggtggtggtggttag	973			

RESULT 2
 AAT96816/C
 ID AAT96816 standard; DNA; 2374 BP.
 XX
 AC AAT96816;
 XX
 DT 12-MAR-1998 (first entry)

XX DNA encoding LysG, LysE and ORF3 from Corynebacterium glutamicum.
 XX LysG; LysE; ORF3; lysine transport; regulatory protein; export protein;
 KW Microbial production; amino acid; animal feed additive; ds.
 XX Corynebacterium glutamicum.
 OS
 XX
 FH Key Location/Qualifiers
 FT CDS complement (82..954)
 FT /*tag= a
 FT /label= LysG
 FT 1016..1726
 FT /*tag= b
 FT /label= LysE
 FT complement (1723..2373)
 FT /*tag= c
 FT /label= orf3
 XX
 PN DEL9548222-A1.
 XX
 PD 26-JUN-1997.
 XX
 XX 22-DEC-1995; 95DE-1048222.
 XX
 PR 22-DEC-1995; 95DE-1048222.
 XX
 PA (KERJ) FORSCHUNGSZENTRUM JUELICH GMBH.
 XX
 XX Eggeling L, Sahm H, Vrljic M;
 PI
 XX WFI; 1997-333867/31.
 DR P-PSDB; AAW37714-16.
 DR
 XX
 PT Increasing microbial production of amino acids, especially lysine -
 PT by improving export carrier activity or corresponding gene
 PT expression, also new export and regulatory genes from
 PT Corynebacterium
 XX
 XX Claim 23 and 26; Page -: 16pp; German.
 PS
 XX
 CC This DNA, isolated from Corynebacterium glutamicum, contains the LysG,
 CC LysE and ORF3 genes. LysG and LysE encode a lysine transport regulatory
 CC protein and an export protein, respectively. Microbial production of
 CC amino acids (A) is improved by increasing the export-carrier activity
 CC and/or the export gene expression in a microorganism that produces (A).
 CC The method is specifically used to increase production of lysine,
 CC used as an animal feed additive. Other (A) are variously useful as
 CC pharmaceuticals, condiments and intermediates for fine chemicals.
 CC This method increases the amount of (A) secreted into the culture medium.
 CC Export of (A) has been found to depend on a single gene.
 CC NB. This sequence has been created from the information given in table 2
 CC of the specification.
 XX
 SQ Sequence 2374 BP; 526 A; 640 C; 648 G; 560 T; 0 other;

Query Match 100.0%; Score 873; DB 18; Length 2374;
 Best Local Similarity 100.0%; Pred. No. 2.9e-268;
 Matches 873; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 atgaacccattcaactggacactttgctctcaatcattgatgaagcagcttcgaaggc 60
 |||
 Db 954 ATGAACCCATTCACTGGACATTTGCTCTCAATCATTTGATGAAGCAGCTTCGAAGGC 895
 |||
 Qy 61 gctccttagccctttccattccctcgcgggtgagtcagcgcgttaagctctcag 120
 |||
 Db 894 GCCTCCTTAGCCCTTTCCATTTCCCTCGCGGTGAGTCAGCGCGTTAAAGCTCTCGAG 835
 |||
 Qy 121 catcacgtggtcagtggttggtatcgcacacccgacccaaagcaaccgaagcgggt 180
 |||
 Db 834 CATCACGTGGTGAGTGTGTTATCGCGACCCCAACCGGCCAAAGCAACCGAAGCGGGT 775

Qy 181 gaadtcttctgcaagcagcgcggaaaaataggtgtgtgctgcaagcagaactaaagcga 240
 |||
 Db 774 GAATCTCTTGTGCAAGCAGCGCGGAAATGGTGTGTGCAAGCAGAACTAAAGCGCAA 715
 |||
 Qy 241 ctatctgagcgccttgcgaaatcccgtaaccatcccatcaacgcagatctcgtatcc 300
 |||
 Db 714 CTATCTGGACGCTTGTGAAATCCCGTTAACCATGCCATCAACGCAGATTCCTATCC 655
 |||
 Qy 301 acatggttctctccctggttcaacgaggtagcttcttggggtggagcaacgcctcacctg 360
 |||
 Db 654 ACATGGTTTCTCTCCGTTCAACGAGGTAGTCTTCTGGGGTGGAGCAACGCTCACGCTG 595
 |||
 Qy 361 cgttggaagatgaagcgcacacattatccttgcgtgcggcggtgagatgttttaggagcg 420
 |||
 Db 594 CGCTTGAAGATGAAGCGCACACATTATCTCTGCTGGCGGTGGAGATGTTTTAGGAGCG 535
 |||
 Qy 421 gtaacccgtgaagctaatcccggtgcgggtagtgaagttagtagaacttggaaacctgcgc 480
 |||
 Db 534 GTAACCGGTGAAGCTAATCCGCTGGCGGGATGTCGAAGTAGTAGAACTTGGAAACCATGGC 475
 |||
 Qy 481 caettggccattgcaacccctcattgcggggtgcctacatggttgatgggaaactagat 540
 |||
 Db 474 CACTTGGCCATTGCAACCCCTCATTTGGGGATGCCTACATGTTGATGGGAAACTAGAT 415
 |||
 Qy 541 tgggctcgatgcgcgtcttacgcttcggtcccaaatgcttcaagaccgtgacctg 600
 |||
 Db 414 TGGGCTCGATGCGCGTCTTACGCTTCGGTCCCAAGATGTGCTCAAGACCGTGACCTG 355
 |||
 Qy 601 gacgggcgcgtcgatggtcctgtgggcgcagcgcgctatccattgtcccgctgcgcgaa 660
 |||
 Db 354 GACGGCGCGTCGATGTCCTGTGGGGCGCAGGCGCGTATCCATTGTCCTCGCGCGGAA 295
 |||
 Qy 661 ggtttggtgaggaattcgcggagcgttgggtgggtgggaacttctccgaacccaagct 720
 |||
 Db 294 GGTTTTGTGAGGCAATTCCGCCGAGGCTTGGTTGGGGACTTCTTCCGAAACCCAAAGCT 235
 |||
 Qy 721 gctccatgctaaaagcagagagaagtatcctcctcgtcgtatgagataccattgacacaccg 780
 |||
 Db 234 GCTCCCATGCTAAAGCAGGAGAGTAGTATCCTCTCGATGAGATACCCATTTGACACACCG 175
 |||
 Qy 781 atgtattggaacatgagcgcctggaatctagatctctagctagactcacagacgcctc 840
 |||
 Db 174 ATGTATTGGCAACATGGCGCTGGAAATCTAGATCTCTTAGCTAGACTACAGACCGCGTC 115
 |||
 Qy 841 gttgatcagcaatcaggggatttcgcgccttag 873
 |||
 Db 114 GTTCATGCAAGCAATCGAGGAGATTGCGGCTTAG 82

RESULT 3
 AAT33536
 ID AAT33536 standard; DNA: 15239 BP.
 XX
 AC -AAT33536;
 XX
 DT 15-FEB-1998 (first entry)
 XX
 DE BCG deletion region 2 and flanking sequences.
 XX
 KW BCG delta 2; virulence; avirulence; attenuation; gene deletion;
 KW mycobacteria; vaccine; infection; marker; ss.
 XX
 OS Mycobacterium bovis strain BCG.
 XX
 FH Key Location/Qualifiers
 FT misc_feature 3382..14071
 FT /*tag= a
 FT /note= "BCG delta 1 deletion region"
 XX
 PN W09625519-A1.
 XX
 XX 22-AUG-1996.
 XX

```

PF 15-FEB-1996; 96WO-US01938.
XX
PR 17-FEB-1995; 95US-0390878.
XX
PA (PATH-) PATHOGENESIS CORP.
XX
PI Mahalras GG, Stover CK;
XX
DR WPI; 1996-393419/39.
XX
PT Detecting markers for avirulence in Mycobacterium - used in
PT production of vaccines against bacterial infection, and to detect
PT bacterial infection
XX
PS Example 1; Fig 2; 66pp; English.
XX
CC This DNA sequence comprises Mycobacterium bovis BCG deletion
CC sequence BCGdelta2. A specific genetic deletion of this region
CC results in an avirulence phenotype of the mycobacterium. 2 Other
CC deletion regions (see AAT33535 and AAT33537) have also been detected.
CC Identification involved screening a BCG cosmid library with a
CC radiolabeled probe obtained following DNA subtraction between
CC virulent Mycobacterium tuberculosis H37Rv and avirulent BCG.
CC The deletions provide useful markers for the identification of an
CC avirulent, or a virulent, mycobacterial phenotype. Determination
CC of avirulence requires the detection of the presence or absence of
CC the deletion; the deletions are detected either by detecting the
CC presence or absence of deletion junctions (see AAT33538-46), or by
CC detecting the presence or absence of the sequences contained within
CC the deletion. Deletion polypeptides are used as components of
CC immunological assays and in vaccines.
XX
SQ Sequence 15239 BP; 2878 A; 4791 C; 4729 G; 2841 T; 0 other;

Query Match 14.0%; Score 122.6; DB 17; Length 15239;
Best Local Similarity 49.9%; Pred. No. 1.7e-28;
Matches 426; Conservative 0; Mismatches 409; Indels 18; Gaps 4;

QY 13 caactggacacttgcctcaatcattgaagcagcttcgaagcgcctcctcttaggc 72
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 5253 cagctggcgcattgctccgtggtcgaactggcagcttcgattgctggcgcgagcgc 5312
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 73 ctttcatttccccctggcgttgagtcgagcgtttaagctctcgagcatcagctgggt 132
DB || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 5313 ctacatgtcacccctcgctgctcagtcagcgcatacgaatgcttgagcagctcggc 5372
QY 133 cgaagtgttgatcgcgaccccaaccgcccgaagcaaccggaagcgggtgaagtctgtg 192
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 5373 cagggtgctggtgctagggaagccatgctcgggcgacgaccgaggtatcccgctgtg 5432
QY 193 caagcagcgcggaaatggtgttgctgcaagcagaaactaaagcgcgaactatctgg--ac 250
DB || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 5433 caggttgccgcgaacacagctgtgctgagtcgagcgcgctgctgaaatgggtggcaac 5492
QY 251 gcttgctgtaa----atccgctaaaccatcgccataacgcagcagattcgctatccacatgg 306
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 5493 gcgtctgaaacgcgcgagcggatcaccattgctggttaaacgcgattccatggcgacatgg 5552
QY 307 tttctcccgctgtcaacgaggtagcttcttgggtggagcaacgctcagctgcgcttg 366
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 5553 tttcggcgtgttcgacg-----gtctcggcgactgctcgtcagcttcggatc 5603
QY 367 gaagtgaagcgcacacattatcttctgctgcgctggagatgttttagagcggtaaac 426
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 5604 gaggaccagaccattccgcgcgctgctacgcggaggtgtgctgattggcgcggtgacc 5663
QY 427 cgtgaagctaaacctggcgggattgtagtagtaagaactggaaacctggaacctgacacttg 486
DB || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 5664 accgagcgaaccccggtgcggggtgcegggtgcacccgctgggtgaaatgcgtacacta 5723
QY 487 gcatgtgaacccctcattgcggagtgctacatggttgatgggaactagatgggct 546
DB || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

RESULT 4

AAF28535/c

ID AAF28535 standard; DNA; 45613 BP.

XX

XX AAF28535;

XX

XX 04-APR-2001 (first entry)

XX

XX Genomic fragment #22.

XX

XX Genomic library; bacteria; human upper airway; otitis media; sinusitis;

KW bronchopulmonary; endocarditis; meningitis; ss.

XX

XX Moraxella catarrhalis.

XX

XX WO200078968-A2.

XX

XX 28-DEC-2000.

XX

XX 16-JUN-2000; 2000WO-US16649.

XX

XX 18-JUN-1999; 99US-0140121.

XX

XX (INCY-) INCYTE GENOMICS INC.

XX

XX Lagace RE, Patterson C, Berg KL;

XX

XX WPI; 2001-041427/05.

XX

XX Genomic library for identifying diagnostic and therapeutic

PT compositions, and for identifying virulence factors, regulatory

PT elements and drug targets, comprises Moraxella catarrhalis nucleic

PT acids.

XX

XX Claim 1; Page 180-191; 545pp; English.

XX

XX The present invention relates to a Moraxella catarrhalis genomic library

CC comprising a combination of 41 nucleic acid molecules (see

CC AAF28514-AAF28554). The library has a number of uses described in the

CC specification e.g. is useful for identifying diagnostic and therapeutic

CC compositions. M. catarrhalis (Branhamella catarrhalis) is a large

CC aerobic, gram-negative diplococcus, normally found among the bacterial

CC flora of human upper airways. M. catarrhalis is known to cause acute,

CC localised infections such as otitis media, sinusitis and bronchopulmonary

CC infection and life-threatening, systemic diseases including endocarditis

CC and meningitis.

Query Match	4.8%	Score 41.6;	DB 21;	Length 2842;
Best Local Similarity	50.5%;	Pred. No. 0.0054;		
Matches 101;	Conservative 0;	Mismatches 99;	Indels 0;	Gaps 0;
Qy 565	ttcgggtcccaagaatgtgtcttcaagacccgtgacctggacgcgcgctgcgatggctcgtg	624		
Db 982	ttcgtccgtgtcgtcgctcctcggtgccgatctgctgccagccgctcgatgatcaggaa	923		
Qy 625	ggggcgagggcgctatccatttcccgtcgcggaagtttttgtagggcaattccgcga	684		
Db 922	cgcgggcgcgcgccgccatttggcgaggctcggaggccctctgtacgccgctcgccggt	863		
Qy 685	ggccttggttggggactcttcccgaaacccaagctgctcccatgctctaaagcaggagaa	744		
Db 862	cagctcggtgtagaaacggcgctcccatgtgcacaccccgaaaccaatccaaagaggacgc	803		
Qy 745	gtgatctctctcgtatgagat	764		
Db 802	cttcttcttcttgcctgtggt	783		

RESULT	7
AAN80093/c	
ID	AAN80093 standard; DNA; 2058 BP.
XX	
AC	
AAN80093;	
XX	
DT	30-MAR-1992 (first entry)
XX	
DE	Sequence of BamHI/SalI fragment of the tfdA gene which encodes 2,4-dichlorophenoxyacetic acid (2,4-D) mono-oxygenase.
DE	
XX	
KW	Herbicide resistance; halogenated aromatic cpd; enzyme;
KW	mono-oxygenase; ss.
XX	
OS	Alcaligenes eutrophus JMP134.
XX	
PH	Key Location/Qualifiers
FT	CDS 748..1608
FT	/*tag= a
FT	/note= "DNA SQ claimed"
XX	
PN	DE3629890-A.
XX	
PD	10-MAR-1988.
XX	
PF	29-AUG-1986; 86DE-3629890.
XX	
PR	29-AUG-1986; 86DE-3629890.
XX	
PA	(SCHD) SCHERING AG.
PA	(STRE/) STREIBER W R.
PA	(MACQ-) MACQUARIE UNIV.
XX	
PI	Streiber W, Timmis KN, Zenk M;
XX	
DR	WPI; 1988-071716/11.
DR	P-PSDB; AAP80079.
XX	
PT	Plasmids and bacteria contg. gene tfdA for 2,4-D-mono:oxygenase -
PT	isolated using new Alcaligenes eutrophus transposon mutants
XX	
PS	Disclosure; Fig 10; 32pp; German.
XX	
CC	The tfdA gene product is a 2,4-D decomposing protein. A.eutrophus
CC	JMP 134 contains an 80 kb plasmid, pJP4, which carries the genes
CC	coding for 2,4-D decomposition. Plasmids and bacteria contg. the
CC	gene are useful for prodn. of 2,4-D mono-oxygenase and for
CC	transferring the ability to decompose 2,4-D to other micro-organisms
CC	(which could then be useful for cleaning waste water polluted by
CC	halogenated aromatic cpds.) or to plants (to impart tolerance to 2,4-
XX	D).
XX	

5Q	Sequence 2058 BP; 372 A; 659 C; 661 G; 366 T; 0 other;
	Query Match 4.2%; Score 37; DB 9; Length 2058; Best Local Similarity 53.9%; pred. No. 0.14; Matches 76; Conservative 0; Mismatches 65; Indels 0; Gaps 0;
Qy	13 caactggacactttgctcctaattgataaagcagcttcgaaggcgctctcttagcc 72
Db	530 CAGCTTCGCTATTTCGTTCGCCGCGGAGGAGGCAACGTCGGTGCCGCCGCGCGCGG 471
Qy	73 ctttcatttccccctcggcggtgagtcagcgcggttaaagctctcgaagcatcaogtgggt 132
Db	470 CTGCATATTTCCAGCCCCCGGTACGCGAGAGATTTCACGGGCTCGAAACAGCATCTGGGC 411
Qy	133 cgagtggttggtatcgcgcaac 153
Db	410 GTGTGTGTGTTCTGAGCGCAGC 390

RESULT	8	
ID	AAQ97857/c	
XX	AAQ97857 standard; DNA; 2058 BP.	
XX		
XX	AAQ97857;	
XX		
DT	06-DEC-1995 (first entry)	
XX		
DE	P. aeruginosa tfdA gene.	
XX		
XX	Transgenic plant; tfdA gene; 2,4-dichlorophenoxyacetic acid; 2,4-D;	
KW	herbicide resistance; sweetgum; Liquidambar styraciflua; hardwood;	
KW	plantation; crop improvement; selectable marker; pUCW200; vector;	
KW	ss.	
XX		
OS	Pseudomonas aeruginosa.	
XX		
FH	Key	Location/Qualifiers
FT	CDS	751..1611
FT		/*tag= a
XX		
PN	W09518862-A.	
XX		
PD	13-JUL-1995.	
XX		
PF	10-JAN-1995; 95WO-US00284.	
XX		
PR	15-DEC-1994; 94US-0358117.	
PR	11-JAN-1994; 94US-0179667.	
XX		
PA	(UNBC) UNION CAMP CORP.	
XX		
PI	Kaphammer BJ;	
XX		
DR	WPI: 1995-255064/33.	
XX	P-PSDB; AAR79659.	
XX		
PT	Using presence of tfdA gene to select transgenic plants - imparting	
PT	resistance to 2,4-di:chloro:phenoxy:acetic acid, esp. sweetgum	
PT	(Liquidambar) trees contg. this gene and plantations of them	
XX		
PS	Disclosure; Fig.7; 56pp; English.	
XX		
CC	A tfdA gene of P. aeruginosa plasmid pRO101, which encodes the	
CC	enzymes for 2,4-D degradation, was cloned into Agrobacterium	
CC	tumefaciens vector pBil121 to form pUCW200. This was used to	
CC	introduce 2,4-D herbicide-resistance into sweetgum, thereby	
CC	allowing selection of transgenic plants and reducing the cost of	
CC	site prepn. and maintenance.	
XX		
SQ	Sequence 2058 BP; 372 A; 659 C; 661 G; 366 T; 0 other;	

Using presence of *tfdA* gene to select transgenic plants - imparting resistance to 2,4-di:chloro:phenoxy:acetic acid, esp. sweetgum (Liquidambar) trees contg. this gene and plantations of them

Disclosure; Fig.7; 56pp; English.

A *tfdA* gene of *P. aeruginosa* plasmid pRO101, which encodes the enzymes for 2,4-D degradation, was cloned into *Agrobacterium tumefaciens* vector pRi121 to form pUCW200. This was used to introduce 2,4-D herbicide-resistance into sweetgum, thereby allowing selection of transgenic plants and reducing the cost of site prepn. and maintenance.

Sequence 2058 BP; 372 A; 659 C; 661 G; 366 T; 0 other;

CC particularly organism-specific probes. Attempts to make efficacious
CC Meningococcus B vaccines have failed mainly due to antigen tolerance.
CC Multivalent vaccines have also been tried but none have successfully
CC overcome antigenic variability. The provision of further, complete
CC sequences may provide an opportunity to identify secreted or surface
CC exposed proteins that may be presumed targets for the immune system and
CC which are not antigenically variable or at least more conserved than
CC other more variable regions.

XX

SQ Sequence 102634 BP; 23871 A; 24828 C; 27888 G; 26042 T; 5 other;

Query Match 4.2%; Score 36.8; DB 21; Length 102634;
Best Local Similarity 52.6%; Pred. No. 0.99;
Matches 80; Conservative 0; Mismatches 72; Indels 0; Gaps

QY 1 atgaaccattcaactggacagcacitctgtcctaatacatgtaagaaggcagcttcgaagcg 60
||| || | ||| | ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 62462 atgatgcgcgtacaattaataatcatttggccgtcgcgacgagggcaaccttacccaa 62521
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 61 gctccttagccctttccaattccccctcgcggtgagtcagcgcgftaaaagtctctcgag 120
||| || | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Db 62522 gccgccaaacgactttctcttcccagcctgcggttctgccccaaattaaagcccttgaa 62581
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 121 catcagctgggtcgagtggttatcgcgcac 152
| | | | | | | | | | | | | | | | | | | | | | | | | | | |

Db 62582 gaatatgtcggcacgccgctgttcaggcgccac 62613
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 10
AAF21609
ID AAF21609 standard; DNA; 349980 BP.
AC AAF21609;
XX
XX
DT 13-MAR-2001 (first entry)
XX
DE Neisseria meningitidis B nucleotide sequence SEQ ID NO:110.
XX
KW Neisseria meningitidis; Neisseria gonorrhoeae; immunogenic; vaccine;
KW diagnosis; antigen; detection; infection; gene therapy; antibacterial;
KW ds.
XX
OS Neisseria meningitidis.
PA
PN WO2000066791-A1.
XX
PD 09-NOV-2000.
XX
PF 08-MAR-2000; 2000WO-US05928.
XX
PR 30-APR-1999; 99US-0132068.
PP 08-OCT-1999; 99WO-US23573.
PR 28-FEB-2000; 2000GB-0004695.
XX
PP (CHIR) CHIRON CORP.
PA
PA (GENO-) INST GENOMIC RES.
XX
PI Pizza M, Hickey E, Peterson J, Tettelin H, Venter JC, Masignani V;
PI Galeotti C, Mora M, Ratti G, Scarcelli M, Scariato V, Rappulli R,
PI Frazer CM, Grandi G;
XX
DR WPI; 2000-647603/62.
XX
XX Neisseria meningitidis B full length genome sequence and open reading
PT frames are used to detect, treat and prevent Neisserial infections -
XX
PS Claim 7; Appendix A; 692pp; English.
XX
XX The present invention describes the full length genome of
CC Neisseria meningitidis B (NMB). The sequences in AAF21544 and AAF21607
CC to AAF21613 represent fragments of the NMB genomic sequence, as the
CC sequence was too long to go in a record on its own it was split into 8
CC

XX

27-1089

Narbonolide synthase; polyketide synthase gene; narbonolide polyketide; antibiotic; C12-hydroxylase; pik; desosamine biosynthesists; desosaminyl transferase enzyme; ketolide; beta-glucosidase enzyme; picromycin biosynthesis; ss.

27-MAY-1999; 99US-0320878.

DE Recombinant cosmid pKOS023-27 containing S. venezuelae PKS genes.

XX Narbonolide polyketide synthase; PKS; cosmid pKOS023-27; ketolide;

KW antibiotic production; narbomycin; picromycin; ds.

XX Streptomyces venezuelae.

XX Key Location/Qualifiers

FT CDS 70..13725

FT /tag= a

FT /product= PICAI

FT /note= "Narbonolide synthase subunit 1"

FT CDS 13830..25049

FT /tag= b

FT /product= PICAI

FT /note= "Narbonolide synthase subunit 2"

FT CDS 25133..29821

FT /tag= c

FT /product= PICAI

FT /note= "Narbonolide synthase subunit 3"

FT CDS 29924..33964

FT /tag= d

FT /product= PICAI

FT /note= "Narbonolide synthase subunit 4"

FT CDS 33961..34806

FT /tag= e

FT /product= PicB

FT /note= "Contains typeII thioesterase domain"

FT CDS 34863..36011

FT /tag= f

FT /product= PICII

FT /note= "4-keto-6-deoxyglucose isomerase"

FT CDS 36159..37439

FT /tag= g

FT /product= PICII

FT /note= "Desosaminyl transferase"

FT CDS 37529..38242

FT /tag= h

FT /product= PICVI

FT /note= "3-amino dimethyltransferase"

XX WO9961599-A2.

XX 02-DEC-1999.

XX 27-MAY-1999; 99WO-US11814.

XX 28-MAY-1998; 98US-0087080.

XX 28-AUG-1998; 98US-0141908.

XX 22-SEP-1998; 98US-0100880.

XX 08-FEB-1999; 99US-0119139.

XX (KOSA-) KOSAN BIOSCIENCES INC.

XX Ashley G, Betlach MC, Betlach M, McDaniel R, Tang L;

XX WPI; 2000-072618/06.

DR P-PSDB; AAY67201, AAY67202, AAY67203, AAY67204, AAY67205, AAY67207,

DR AAY67208, AAY67211.

XX New recombinant DNA encoding a domain of narbonolide polyketide synthase, for production of ketolide antibiotics -

XX Example 2; Page 16-27; 98pp; English.

XX This is the recombinant cosmid pKOS023-27 DNA sequence which contains a Streptomyces venezuelae DNA insert. The cosmid contains open reading frames which encode the various modules of the narbonolide polyketide synthase (PKS). The invention relates to recombinant DNA containing a coding sequence for a narbonolide PKS. Polyketides are compounds synthesised from 2-carbon units through a series of condensations and subsequent modifications. Modular PKSS are responsible for the production of many antibiotics including picromycin. The narbonolide PKS consists of

CC a loading module, six extender modules, and two thioester domains. Four proteins make up the narbonolide PKS (PICAI, PICAI, PICAI and PICAI).

CC PICAI includes the loading module and extender modules 1 and 2, PICAI includes extender modules 3 and 4, PICAI includes extender module 5 and PICAI includes extender module 6 and a type II thioesterase domain. The second type II thioesterase domain is found on the PICB protein. The nucleotide sequences encoding all of these proteins can be isolated in recombinant form from the recombinant cosmid pKOS023-27. Narbonolide is desaminylated in S. venezuelae to yield narbomycin, and the desosamine transferase enzyme is required for this conversion, and the desosamine biosynthetic genes are also found in cosmid pKOS023-27. The recombinant DNA of the invention is used to express, in transformed cells, narbonolide (or its derivatives) or other ketolides (particularly hybrids), which may then be converted (e.g. by other enzymes recombinantly expressed in the same hosts) to polyketide antibiotics or their intermediates. The antibiotics are useful in human or veterinary medicine.

XX

SQ Sequence 38506 BP; 4914 A; 15118 C; 13444 G; 5030 T; 0 other;

Query Match 4.0%; Score 35; DB 21; Length 38506;

Best Local Similarity 45.7%; Pred. No. 2.3;

Matches 122; Conservative 0; Mismatches 145; Indels 0; Gaps 0;

Qy 406 gatgttttaggcggttaacccgtgaagctaatccctggtggcgatgtgaagtagtagaa 465

Db 28123 GAGGTCTGTCGAGGAGCAGACGCCGCTGAAGCGCTGCGCGGTGAG 28064

Qy 466 ctggaaccatgcaccattgcccattcaacccctcatcggtggtgctacatggtt 525

Db 28063 CCGGCGCGGAGGCGCTCACGGTCTGTCGCCGCCCGCTTCCAGTACGTGACCTCCGC 28004

Qy 526 gatggaaactagattgggtgcgatgccgtcttaacgtcttcggtcccaagatgtgctt 585

Db 28003 TCCGCGGCGCTTACGCGGCCAGCAGCTCGCTTCTGCTCGCTCCCGCGGCGAGAT 27944

Qy 586 caagaccgtgacctggacggcggtgcgtgcgttcctgtggggcgagcgcgatccatt 645

Db 27943 CCACGCCCGGACAGGTCGCGCTGCGCGGAGCGGCTGACGCGCTTCCACTCGACGCGGTA 27884

Qy 646 gtcccgctgcgggaaggttttggtag 672

Db 27883 ACGCCAGTCGTCGCGCGCGCTGGTGG 27857

Search completed: September 14, 2001, 07:58:58

Job time: 9147 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 14, 2001, 06:59:12 ; Search time 1873.17 Seconds
(without alignments)
4405.548 Million cell updates/sec

Title: US-09-105-117I-1-COPY_1421_2293
Perfect score: 873
Sequence: 1 atgaacccattcaactgga.....tcgagggttgccgcttag 873

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0
Searched: 10228115 seqs, 4726426750 residues

Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- 1: gb_est1.*
- 2: gb_est2.*
- 3: gb_est3.*
- 4: gb_est4.*
- 5: gb_est5.*
- 6: gb_est6.*
- 7: gb_est7.*
- 8: gb_est8.*
- 9: gb_est9.*
- 10: gb_est10.*
- 11: gb_est11.*
- 12: gb_est12.*
- 13: gb_est13.*
- 14: gb_est14.*
- 15: gb_est15.*
- 16: gb_est16.*
- 17: gb_est17.*
- 18: gb_est18.*
- 19: gb_est19.*
- 20: gb_est20.*
- 21: gb_est21.*
- 22: gb_est22.*
- 23: gb_est23.*
- 24: gb_est24.*
- 25: gb_est25.*
- 26: gb_est26.*
- 27: gb_est27.*
- 28: gb_est28.*
- 29: gb_est29.*
- 30: gb_est30.*
- 31: gb_est31.*
- 32: gb_est32.*
- 33: gb_est33.*
- 34: gb_est34.*
- 35: gb_est35.*
- 36: gb_est36.*
- 37: gb_est37.*
- 38: gb_est38.*
- 39: gb_est39.*
- 40: gb_est40.*
- 41: em_est1.*
- 42: em_est2.*
- 43: em_est3.*
- 44: em_est4.*
- 45: em_est5.*
- 46: em_est6.*
- 47: em_est7.*
- 48: em_est8.*
- 49: em_est9.*
- 50: em_est10.*
- 51: em_est11.*
- 52: em_est12.*
- 53: em_est13.*
- 54: em_est14.*
- 55: em_est15.*
- 56: em_est16.*
- 57: em_est17.*
- 58: em_est18.*
- 59: em_est19.*
- 60: em_est20.*
- 61: em_est21.*
- 62: em_est22.*
- 63: em_est23.*
- 64: em_est24.*
- 65: em_est25.*
- 66: em_est26.*
- 67: em_est27.*
- 68: em_est28.*
- 69: em_est29.*
- 70: em_est30.*
- 71: em_est31.*
- 72: em_est32.*
- 73: em_est33.*
- 74: em_est34.*
- 75: em_est35.*
- 76: em_est36.*
- 77: em_est37.*
- 78: em_est38.*
- 79: em_est39.*
- 80: em_est40.*
- 81: em_est41.*
- 82: em_est42.*
- 83: em_est43.*
- 84: em_est44.*
- 85: em_est45.*
- 86: em_est46.*
- 87: em_est47.*
- 88: em_est48.*
- 89: em_est49.*
- 90: em_est50.*
- 91: em_est51.*
- 92: em_est52.*
- 93: em_est53.*
- 94: em_est54.*
- 95: em_est55.*
- 96: em_est56.*
- 97: em_est57.*
- 98: em_est58.*
- 99: em_est59.*
- 100: em_est60.*
- 101: em_est61.*
- 102: em_est62.*
- 103: em_est63.*
- 104: em_est64.*
- 105: em_est65.*
- 106: em_est66.*
- 107: em_est67.*
- 108: em_est68.*
- 109: em_est69.*
- 110: em_est70.*
- 111: em_est71.*
- 112: em_est72.*
- 113: em_est73.*
- 114: em_est74.*
- 115: em_est75.*
- 116: em_est76.*
- 117: em_est77.*
- 118: em_est78.*
- 119: em_est79.*
- 120: em_est80.*
- 121: em_est81.*
- 122: em_est82.*
- 123: em_est83.*
- 124: em_est84.*
- 125: em_est85.*
- 126: em_est86.*
- 127: em_est87.*
- 128: em_est88.*
- 129: em_est89.*
- 130: em_est90.*
- 131: em_est91.*
- 132: em_est92.*
- 133: em_est93.*
- 134: em_est94.*
- 135: em_est95.*
- 136: em_est96.*
- 137: em_est97.*
- 138: em_est98.*
- 139: em_est99.*
- 140: em_est100.*
- 141: em_est101.*
- 142: em_est102.*
- 143: em_est103.*
- 144: em_est104.*
- 145: em_est105.*
- 146: em_est106.*
- 147: em_est107.*
- 148: em_est108.*
- 149: em_est109.*
- 150: em_est110.*
- 151: em_est111.*
- 152: em_est112.*
- 153: em_est113.*
- 154: em_est114.*
- 155: em_est115.*
- 156: em_est116.*
- 157: em_est117.*
- 158: em_est118.*
- 159: em_est119.*
- 160: em_est120.*
- 161: em_est121.*
- 162: em_est122.*
- 163: em_est123.*
- 164: em_est124.*
- 165: em_est125.*
- 166: em_est126.*
- 167: em_est127.*
- 168: em_est128.*
- 169: em_est129.*
- 170: em_est130.*
- 171: em_est131.*
- 172: em_est132.*
- 173: em_est133.*
- 174: em_est134.*
- 175: em_est135.*
- 176: em_est136.*
- 177: em_est137.*
- 178: em_est138.*
- 179: em_est139.*
- 180: em_est140.*
- 181: em_est141.*
- 182: em_est142.*
- 183: em_est143.*
- 184: em_est144.*
- 185: em_est145.*
- 186: em_est146.*
- 187: em_est147.*
- 188: em_est148.*
- 189: em_est149.*
- 190: em_est150.*
- 191: em_est151.*
- 192: em_est152.*
- 193: em_est153.*
- 194: em_est154.*
- 195: em_est155.*
- 196: em_est156.*
- 197: em_est157.*
- 198: em_est158.*
- 199: em_est159.*
- 200: em_est160.*
- 201: em_est161.*
- 202: em_est162.*
- 203: em_est163.*
- 204: em_est164.*
- 205: em_est165.*
- 206: em_est166.*
- 207: em_est167.*
- 208: em_est168.*
- 209: em_est169.*
- 210: em_est170.*
- 211: em_est171.*
- 212: em_est172.*
- 213: em_est173.*
- 214: em_est174.*
- 215: em_est175.*
- 216: em_est176.*
- 217: em_est177.*
- 218: em_est178.*
- 219: em_est179.*
- 220: em_est180.*
- 221: em_est181.*
- 222: em_est182.*
- 223: em_est183.*
- 224: em_est184.*
- 225: em_est185.*
- 226: em_est186.*
- 227: em_est187.*
- 228: em_est188.*
- 229: em_est189.*
- 230: em_est190.*
- 231: em_est191.*
- 232: em_est192.*
- 233: em_est193.*
- 234: em_est194.*
- 235: em_est195.*
- 236: em_est196.*
- 237: em_est197.*
- 238: em_est198.*
- 239: em_est199.*
- 240: em_est200.*
- 241: em_est201.*
- 242: em_est202.*
- 243: em_est203.*
- 244: em_est204.*
- 245: em_est205.*
- 246: em_est206.*
- 247: em_est207.*
- 248: em_est208.*
- 249: em_est209.*
- 250: em_est210.*
- 251: em_est211.*
- 252: em_est212.*
- 253: em_est213.*
- 254: em_est214.*
- 255: em_est215.*
- 256: em_est216.*
- 257: em_est217.*
- 258: em_est218.*
- 259: em_est219.*
- 260: em_est220.*
- 261: em_est221.*
- 262: em_est222.*
- 263: em_est223.*
- 264: em_est224.*
- 265: em_est225.*
- 266: em_est226.*
- 267: em_est227.*
- 268: em_est228.*
- 269: em_est229.*
- 270: em_est230.*
- 271: em_est231.*
- 272: em_est232.*
- 273: em_est233.*
- 274: em_est234.*
- 275: em_est235.*
- 276: em_est236.*
- 277: em_est237.*
- 278: em_est238.*
- 279: em_est239.*
- 280: em_est240.*
- 281: em_est241.*
- 282: em_est242.*
- 283: em_est243.*
- 284: em_est244.*
- 285: em_est245.*
- 286: em_est246.*
- 287: em_est247.*
- 288: em_est248.*
- 289: em_est249.*
- 290: em_est250.*
- 291: em_est251.*
- 292: em_est252.*
- 293: em_est253.*
- 294: em_est254.*
- 295: em_est255.*
- 296: em_est256.*
- 297: em_est257.*
- 298: em_est258.*
- 299: em_est259.*
- 300: em_est260.*
- 301: em_est261.*
- 302: em_est262.*
- 303: em_est263.*
- 304: em_est264.*
- 305: em_est265.*
- 306: em_est266.*
- 307: em_est267.*
- 308: em_est268.*
- 309: em_est269.*
- 310: em_est270.*
- 311: em_est271.*
- 312: em_est272.*
- 313: em_est273.*
- 314: em_est274.*
- 315: em_est275.*
- 316: em_est276.*
- 317: em_est277.*
- 318: em_est278.*
- 319: em_est279.*
- 320: em_est280.*
- 321: em_est281.*
- 322: em_est282.*
- 323: em_est283.*
- 324: em_est284.*
- 325: em_est285.*
- 326: em_est286.*
- 327: em_est287.*
- 328: em_est288.*
- 329: em_est289.*
- 330: em_est290.*
- 331: em_est291.*
- 332: em_est292.*
- 333: em_est293.*
- 334: em_est294.*
- 335: em_est295.*
- 336: em_est296.*
- 337: em_est297.*
- 338: em_est298.*
- 339: em_est299.*
- 340: em_est300.*
- 341: em_est301.*
- 342: em_est302.*
- 343: em_est303.*
- 344: em_est304.*
- 345: em_est305.*
- 346: em_est306.*
- 347: em_est307.*
- 348: em_est308.*
- 349: em_est309.*
- 350: em_est310.*
- 351: em_est311.*
- 352: em_est312.*
- 353: em_est313.*
- 354: em_est314.*
- 355: em_est315.*
- 356: em_est316.*
- 357: em_est317.*
- 358: em_est318.*
- 359: em_est319.*
- 360: em_est320.*
- 361: em_est321.*
- 362: em_est322.*
- 363: em_est323.*
- 364: em_est324.*
- 365: em_est325.*
- 366: em_est326.*
- 367: em_est327.*
- 368: em_est328.*
- 369: em_est329.*
- 370: em_est330.*
- 371: em_est331.*
- 372: em_est332.*
- 373: em_est333.*
- 374: em_est334.*
- 375: em_est335.*
- 376: em_est336.*
- 377: em_est337.*
- 378: em_est338.*
- 379: em_est339.*
- 380: em_est340.*
- 381: em_est341.*
- 382: em_est342.*
- 383: em_est343.*
- 384: em_est344.*
- 385: em_est345.*
- 386: em_est346.*
- 387: em_est347.*
- 388: em_est348.*
- 389: em_est349.*
- 390: em_est350.*
- 391: em_est351.*
- 392: em_est352.*
- 393: em_est353.*
- 394: em_est354.*
- 395: em_est355.*
- 396: em_est356.*
- 397: em_est357.*
- 398: em_est358.*
- 399: em_est359.*
- 400: em_est360.*
- 401: em_est361.*
- 402: em_est362.*
- 403: em_est363.*
- 404: em_est364.*
- 405: em_est365.*
- 406: em_est366.*
- 407: em_est367.*
- 408: em_est368.*
- 409: em_est369.*
- 410: em_est370.*
- 411: em_est371.*
- 412: em_est372.*
- 413: em_est373.*
- 414: em_est374.*
- 415: em_est375.*
- 416: em_est376.*
- 417: em_est377.*
- 418: em_est378.*
- 419: em_est379.*
- 420: em_est380.*
- 421: em_est381.*
- 422: em_est382.*
- 423: em_est383.*
- 424: em_est384.*
- 425: em_est385.*
- 426: em_est386.*
- 427: em_est387.*
- 428: em_est388.*
- 429: em_est389.*
- 430: em_est390.*
- 431: em_est391.*
- 432: em_est392.*
- 433: em_est393.*
- 434: em_est394.*
- 435: em_est395.*
- 436: em_est396.*
- 437: em_est397.*
- 438: em_est398.*
- 439: em_est399.*
- 440: em_est400.*
- 441: em_est401.*
- 442: em_est402.*
- 443: em_est403.*
- 444: em_est404.*
- 445: em_est405.*
- 446: em_est406.*
- 447: em_est407.*
- 448: em_est408.*
- 449: em_est409.*
- 450: em_est410.*
- 451: em_est411.*
- 452: em_est412.*
- 453: em_est413.*
- 454: em_est414.*
- 455: em_est415.*
- 456: em_est416.*
- 457: em_est417.*
- 458: em_est418.*
- 459: em_est419.*
- 460: em_est420.*
- 461: em_est421.*
- 462: em_est422.*
- 463: em_est423.*
- 464: em_est424.*
- 465: em_est425.*
- 466: em_est426.*
- 467: em_est427.*
- 468: em_est428.*
- 469: em_est429.*
- 470: em_est430.*
- 471: em_est431.*
- 472: em_est432.*
- 473: em_est433.*
- 474: em_est434.*
- 475: em_est435.*
- 476: em_est436.*
- 477: em_est437.*
- 478: em_est438.*
- 479: em_est439.*
- 480: em_est440.*
- 481: em_est441.*
- 482: em_est442.*
- 483: em_est443.*
- 484: em_est444.*
- 485: em_est445.*
- 486: em_est446.*
- 487: em_est447.*
- 488: em_est448.*
- 489: em_est449.*
- 490: em_est450.*
- 491: em_est451.*
- 492: em_est452.*
- 493: em_est453.*
- 494: em_est454.*
- 495: em_est455.*
- 496: em_est456.*
- 497: em_est457.*
- 498: em_est458.*
- 499: em_est459.*
- 500: em_est460.*
- 501: em_est461.*
- 502: em_est462.*
- 503: em_est463.*
- 504: em_est464.*
- 505: em_est465.*
- 506: em_est466.*
- 507: em_est467.*
- 508: em_est468.*
- 509: em_est469.*
- 510: em_est470.*
- 511: em_est471.*
- 512: em_est472.*
- 513: em_est473.*
- 514: em_est474.*
- 515: em_est475.*
- 516: em_est476.*
- 517: em_est477.*
- 518: em_est478.*
- 519: em_est479.*
- 520: em_est480.*
- 521: em_est481.*
- 522: em_est482.*
- 523: em_est483.*
- 524: em_est484.*
- 525: em_est485.*
- 526: em_est486.*
- 527: em_est487.*
- 528: em_est488.*
- 529: em_est489.*
- 530: em_est490.*
- 531: em_est491.*
- 532: em_est492.*
- 533: em_est493.*
- 534: em_est494.*
- 535: em_est495.*
- 536: em_est496.*
- 537: em_est497.*
- 538: em_est498.*
- 539: em_est499.*
- 540: em_est500.*
- 541: em_est501.*
- 542: em_est502.*
- 543: em_est503.*
- 544: em_est504.*
- 545: em_est505.*
- 546: em_est506.*
- 547: em_est507.*
- 548: em_est508.*
- 549: em_est509.*
- 550: em_est510.*
- 551: em_est511.*
- 552: em_est512.*
- 553: em_est513.*
- 554: em_est514.*
- 555: em_est515.*
- 556: em_est516.*
- 557: em_est517.*
- 558: em_est518.*
- 559: em_est519.*
- 560: em_est520.*
- 561: em_est521.*
- 562: em_est522.*
- 563: em_est523.*
- 564: em_est524.*
- 565: em_est525.*
- 566: em_est526.*
- 567: em_est527.*
- 568: em_est528.*
- 569: em_est529.*
- 570: em_est530.*
- 571: em_est531.*
- 572: em_est532.*
- 573: em_est533.*
- 574: em_est534.*
- 575: em_est535.*
- 576: em_est536.*
- 577: em_est537.*
- 578: em_est538.*
- 579: em_est539.*
- 580: em_est540.*
- 581: em_est541.*
- 582: em_est542.*
- 583: em_est543.*
- 584: em_est544.*
- 585: em_est545.*
- 586: em_est546.*
- 587: em_est547.*
- 588: em_est548.*
- 589: em_est549.*
- 590: em_est550.*
- 591: em_est551.*
- 592: em_est552.*
- 593: em_est553.*
- 594: em_est554.*
- 595: em_est555.*
- 596: em_est556.*
- 597: em_est557.*
- 598: em_est558.*
- 599: em_est559.*
- 600: em_est560.*
- 601: em_est561.*
- 602: em_est562.*
- 603: em_est563.*
- 604: em_est564.*
- 605: em_est565.*
- 606: em_est566.*
- 607: em_est567.*
- 608: em_est568.*
- 609: em_est569.*
- 610: em_est570.*
- 611: em_est571.*
- 612: em_est572.*
- 613: em_est573.*
- 614: em_est574.*
- 615: em_est575.*
- 616: em_est576.*
- 617: em_est577.*
- 618: em_est578.*
- 619: em_est579.*
- 620: em_est580.*
- 621: em_est581.*
- 622: em_est582.*
- 623: em_est583.*
- 624: em_est584.*
- 625: em_est585.*
- 626: em_est586.*
- 627: em_est587.*
- 628: em_est588.*
- 629: em_est589.*
- 630: em_est590.*
- 631: em_est591.*
- 632: em_est592.*
- 633: em_est593.*
- 634: em_est594.*
- 635: em_est595.*
- 636: em_est596.*
- 637: em_est597.*
- 638: em_est598.*
- 639: em_est599.*
- 640: em_est600.*
- 641: em_est601.*
- 642: em_est602.*
- 643: em_est603.*
- 644: em_est604.*
- 645: em_est605.*
- 646: em_est606.*
- 647: em_est607.*
- 648: em_est608.*
- 649: em_est609.*
- 650: em_est610.*
- 651: em_est611.*
- 652: em_est612.*
- 653: em_est613.*
- 654: em_est614.*
- 655: em_est615.*
- 656: em_est616.*
- 657: em_est617.*
- 658: em_est618.*
- 659: em_est619.*
- 660: em_est620.*
- 661: em_est621.*
- 662: em_est622.*
- 663: em_est623.*
- 664: em_est624.*
- 665: em_est625.*
- 666: em_est626.*
- 667: em_est627.*
- 668: em_est628.*
- 669: em_est629.*
- 670: em_est630.*
- 671: em_est631.*
- 672: em_est632.*
- 673: em_est633.*
- 674: em_est634.*
- 675: em_est635.*
- 676: em_est636.*
- 677: em_est637.*
- 678: em_est638.*
- 679: em_est639.*
- 680: em_est640.*
- 681: em_est641.*
- 682: em_est642.*
- 683: em_est643.*
- 684: em_est644.*
- 685: em_est645.*
- 686: em_est646.*
- 687: em_est647.*
- 688: em_est648.*
- 689: em_est649.*
- 690: em_est650.*
- 691: em_est651.*
- 692: em_est652.*
- 693: em_est653.*
- 694: em_est654.*
- 695: em_est655.*
- 696: em_est656.*
- 697: em_est657.*
- 698: em_est658.*
- 699: em_est659.*
- 700: em_est660.*
- 701: em_est661.*
- 702: em_est662.*
- 703: em_est663.*
- 704: em_est664.*
- 705: em_est665.*
- 706: em_est666.*
- 707: em_est667.*
- 708: em_est668.*
- 709: em_est669.*
- 710: em_est670.*
- 711: em_est671.*
- 712: em_est672.*
- 713: em_est673.*
- 714: em_est674.*
- 715: em_est675.*
- 716: em_est676.*
- 717: em_est677.*
- 718: em_est678.*
- 719: em_est679.*
- 720: em_est680.*
- 721: em_est681.*
- 722: em_est682.*
- 723: em_est683.*
- 724: em_est684.*
- 725: em_est685.*
- 726: em_est686.*
- 727: em_est687.*
- 728: em_est688.*
- 729: em_est689.*
- 730: em_est690.*
- 731: em_est691.*
- 732: em_est692.*
- 733: em_est693.*
- 734: em_est694.*
- 735: em_est695.*
- 736: em_est696.*
- 737: em_est697.*
- 738: em_est698.*
- 739: em_est699.*
- 740: em_est700.*
- 741: em_est701.*
- 742: em_est702.*
- 743: em_est703.*
- 744: em_est704.*
- 745: em_est705.*
- 746: em_est706.*
- 747: em_est707.*
- 748: em_est708.*
- 749: em_est709.*
- 750: em_est710.*
- 751: em_est711.*
- 752: em_est712.*
- 753: em_est713.*
- 754: em_est714.*
- 755: em_est715.*
- 756: em_est716.*
- 757: em_est717.*
- 758: em_est718.*
- 759: em_est719.*
- 760: em_est720.*
- 761: em_est721.*
- 762: em_est722.*
- 763: em_est723.*
- 764: em_est724.*
- 765: em_est725.*
- 766: em_est726.*
- 767: em_est727.*
- 768: em_est728.*
- 769: em_est729.*
- 770: em_est730.*
- 771: em_est731.*
- 772: em_est732.*
- 773: em_est733.*
- 774: em_est734.*
- 775: em_est735.*
- 776: em_est736.*
- 777: em_est737.*
- 778: em_est738.*
- 779: em_est739.*
- 780: em_est740.*
- 781: em_est741.*
- 782: em_est742.*
- 783: em_est743.*
- 784: em_est744.*
- 785: em_est745.*
- 786: em_est746.*
- 787: em_est747.*
- 788: em_est748.*
- 789: em_est749.*
- 790: em_est750.*
- 791: em_est751.*
- 792: em_est752.*
- 793: em_est753.*</

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est67:*
137: gb_est68:*
138: gb_est69:*
139: gb_est70:*
140: gb_est71:*
141: gb_est72:*
142: gb_est73:*
143: gb_est74:*
144: gb_est75:*
145: gb_est76:*
146: gb_est77:*
147: gb_est78:*
148: gb_est79:*
149: gb_est80:*
150: gb_est81:*
151: gb_est82:*
152: gb_est83:*
153: gb_est84:*
154: gb_est85:*
155: gb_est86:*
156: gb_est87:*
157: gb_est88:*
158: gb_est89:*
159: gb_est90:*
160: gb_est91:*
161: gb_est92:*
162: gb_est93:*
163: gb_est94:*
164: gb_est95:*
165: gb_est96:*
166: gb_est97:*
167: gb_est98:*
168: gb_est99:*
169: gb_est100:*
170: gb_est101:*
171: gb_est102:*
172: gb_est103:*
173: gb_est104:*
174: gb_est105:*
175: gb_est106:*
176: gb_est107:*
177: gb_est108:*
178: gb_est109:*
179: gb_est110:*
180: gb_est111:*
181: gb_est112:*
182: gb_est113:*
183: gb_est114:*
184: gb_est115:*
185: gb_est116:*
186: gb_est117:*
187: gb_est118:*
188: gb_est119:*
189: gb_est120:*

190: gb_est121:*
191: gb_est122:*
192: gb_est123:*
193: gb_est124:*
194: gb_est125:*
195: gb_est126:*
196: gb_est127:*
197: gb_est128:*
198: gb_est129:*
199: gb_est130:*
200: gb_est131:*
201: gb_est132:*
202: gb_est133:*
203: gb_est134:*
204: gb_est135:*
205: gb_est136:*
206: gb_est137:*
207: gb_est138:*
208: gb_est139:*
209: gb_est140:*
210: gb_est141:*
211: gb_est142:*
212: gb_est143:*
213: gb_est144:*
214: gb_est145:*
215: gb_est146:*
216: gb_est147:*
217: gb_est148:*
218: gb_est149:*
219: gb_est150:*
220: gb_est151:*
221: gb_est152:*
222: gb_est153:*
223: gb_est154:*
224: gb_est155:*
225: gb_est156:*
226: gb_est157:*
227: gb_est158:*
228: gb_est159:*
229: gb_est160:*
230: gb_est161:*
231: gb_est162:*
232: gb_est163:*
233: gb_est164:*
234: gb_est165:*
235: gb_est166:*
236: gb_est167:*
237: gb_est168:*
238: gb_est169:*
239: gb_est170:*
240: gb_est171:*
241: gb_est172:*
242: gb_est173:*
243: gb_est174:*
244: gb_est175:*
245: gb_est176:*
246: gb_est177:*
247: gb_est178:*
248: gb_est179:*
249: gb_est180:*
250: gb_est181:*
251: gb_est182:*
252: gb_est183:*
253: gb_est184:*
254: gb_est185:*
255: gb_est186:*
256: gb_est187:*
257: gb_est188:*
258: gb_est189:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBelOBAcl1.

FEATURES
source
1. .1101
/organism="Drosophila melanogaster"
/plasmid="pBelOBAcl1"
/db_xref="taxon:7227"
/clone_lib="DrosBAC"
/clone="BACN37108"
/note="end : SP6"

BASE COUNT 254 a 176 c 160 g 152 t 359 others
ORIGIN

Query Match 4.4%; Score 38.8; DB 219; Length 1101;
Best Local Similarity 14.4%; Pred. No. 0.95;
Matches 50; Conservative 153; Mismatches 145; Indels 0; Gaps 0;
QY 350 cgtcagcgtcgctggagatgaagcgcacacattatcttgcgtgcggtggagatg 409
Db 1054 SSSBSBBSGSSSSBSBSSTSSSSSTSSSSSTSSSSTSSSSTTTTTTTT 995
QY 410 ttttagagcgttaaccgtgaagcgttaaccgtggcggtgtagtagtaacttg 469
Db 994 TTSTTTBBSBSBSBSSTSSSSTSSSSTSSSSTSSSSTSSSSTSSSSTTBS 935
QY 470 gaaccatgcacactggcattgaacccctcattggcggtgctacatggttgatg 529
Db 934 TSBBSTSSSSSSSSBSBSSTSSSSTSSSSTSSSSTSSSSTSSSSTSSS 875
QY 530 ggaactagatggctggcgatccgctctacgtctgcgtcccaagaatgcttcaag 589
Db 874 SSSBSBSSTTTSSSTSSSBSBSSTSSSSTSSSSTSSSSTSSSSTSSS 815
QY 590 accgtgacgtgacggcggtcgatggtctggtggcgagcggtatccatgctcc 649
Db 814 SSBGCGSCSTGSCSSBSBCKSTSGSTSSGTTGBCSTGCGCCCTCTGCG 755
QY 650 cgtcgcggaagattgttgaggaatcgcgcagccttgggtgg 697
Db 754 TSTCCSCSGTSSBKCCYSTSSASTGSSCSSTSSSGGTGKMTGG 707

RESULT 3
CN50091P
LOCUS
DEFINITION
Drosophila melanogaster genome survey sequence TET3 end of BAC # BACR19D16 of RPI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
fruit fly.
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
Submitted (02-JUN-1999) Genoscope - Centre National de Sequençage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammosier in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,

NY. The library is named RPI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's pl and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES
Location/Qualifiers
source
1. .925
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="RPI-98"
/clone="BACR19D16"
/note="end : TET3"

BASE COUNT 120 a 61 c 61 g 172 t 511 others
ORIGIN

Query Match 4.3%; Score 37.6; DB 219; Length 925;
Best Local Similarity 15.0%; Pred. No. 2.1;
Matches 48; Conservative 130; Mismatches 142; Indels 0; Gaps 0;
QY 343 ggaagcaactcagcgtcgctggagatgaagcgcacacattatcttgcgtgcg 402
Db 603 SSBSSSKCSTSSBSBSCSSKSVCTSSSSSSSSSSSSSSSSSSSSSSSS 662
QY 403 ggagatgttttaggagcgttaaccgtgaagcgttaaccgtggcggtgtagtagta 462
Db 663 SSSSSSYTTSTKSTASGSGSWAGGGGTGTTSSSSSSSSSSSSSSSSSS 722
QY 463 gaactggaaacatgcgcactggcattggcattgaacccctcattggcggtgctacatg 522
Db 723 GBSSSSGSSSSSTSSBBSCTSTSSSSSSSSSSTCCCTCCSYSYSSSTSS 782
QY 523 gttgtaggaactagatggcgtcgatgcgtcttctacgtctgcgtcccaagaatg 582
Db 783 STSGSSSSSVGTSSSSDSTSTCCSCCYMCTCCTSTBMCYTSTSCGSSSS 842
QY 583 cttcaagacgtgacatgcgcagggcgctcgatggtcctgtggcgagcgctatcc 642
Db 843 CGCGGCGSSSTNGMBGTSSACSSSSSSSSSSSSSSSSSSSSSSSSSSSS 902
QY 643 attgtccctcgcggaag 662
Db 903 ASKSSSSGSSVSGSGSGSGS 922

RESULT 4
BF503196/c
LOCUS
DEFINITION
BF503196 667 bp mRNA EST 16-APR-2001
Drosophila melanogaster cDNA clone AT19123 5 similar to CG11737: FBan0017137 last updated:000321, mRNA sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
fruit fly.
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
Stapleton, M., Brokstein, P., Hong, L., Agbayani, A., Baxter, E., Berman, D., Carlson, J., Chavez, C., Chew, M., Dorsett, V., Farfan, D., Frise, E., George, R., Gonzalez, M., Guarin, H., Harris, N., Li, P., Liao, G., Miranda, A., Misra, S., Mungall, C. J., Nunoo, J., Pacle, J., Park, S., Paragas, V., Phouanavong, S., Wan, K., Yu, C., Lewis, S. E., Celnikier, S. and Rubin, G. M.
Berkeley Drosophila Gene Collection Project
Unpublished (2000)
Contact: Stapleton, M.
BDGP
Lawrence Berkeley National Lab


```

/db_xref="taxon:9606"
/clone="IMAGE:3603764"
/clone_lib="NIH_MGC_44"
/tissue_type="endometrium, adenocarcinoma cell line"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: uterus; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT      173 a   171 c   269 g   172 t
ORIGIN

Query Match      4.1%; Score 36; DB 167; Length 785;
Best Local Similarity 62.08; Pred. No. 6.2; Indels 0; Gaps 0
Matches 57; Conservative 0; Mismatches 35;

QY 604 gggcgctcgatgctctgtggggcgagcgctatccattgtccgcgtcgcggaaggt 663
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 214 GGGCGCTCTCGTCCGTGGGGGCCCTGGCGGGGTATCCATGTTCGGTGGCTGGAGGT 273

QY 664 ttggtgaggaattcgcgagggccttggttg 695
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 274 GCTGGAGGAGGATTCGACAAAGCTTTGGTGG 305

RESULT 6
BF485823      650 bp      mRNA      EST      16-APR-2001
LOCUS      AT19875.5prime AT Drosophila melanogaster adult testes pOTB7
DEFINITION      Drosophila melanogaster cDNA clone AT19875 5 similar to CGI3768:
                  FBan0013768 last_updated:000321, mRNA sequence.
ACCESSION      BF485823
VERSION      BF485823.1 GI:11569124
KEYWORDS      EST.
SOURCE      fruit fly.
ORGANISM      Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 650)
Stapleton,M., Brokstein,P., Hong,L., Agbayani,A., Baxter,E., Berman
,B., Carlson,J., Champe,M., Chavez,C., Chew,M., Dorsett,V., Farfan
,D., Frise,E., George,R., Gonzalez,M., Guarin,H., Harris,N., Li,P.,
Liao,G., Miranda,A., Misra,S., Mungall,C.J., Nunoo,J., Pacleb,J.,
Park,S., Paragas,V., Phouanavong,S., Wan,K., Yu,C., Lewis,S.E.,
celniker,S. and Rubin,G.M.
Berkeley Drosophila Gene Collection Project
Unpublished (2000)
Contact: Stapleton, M.
BDGP
Lawrence Berkeley National Lab
One Cyclotron Rd, Berkeley, CA 94720, USA
Fax: 510 486 6798
Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
hit genomic sequence AE003614
Plate: AT 198 row: G column: 3
High quality sequence stop: 633.
Location/Qualifiers
1. .650
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="AT19875"
/clone_lib="AT Drosophila melanogaster adult testes pOTB7"
/sex="male"
/dev_stage="0-3 day old Ore-R males"
/lab_host="Plates AT.10-AT.120: DH5-alpha. Plates
AT.121-AT.319: DH5-alpha Tona"
/notes="Organ: ADULT testes; Vector: pOTB7; Site_1: EcoRI;
Site_2: XhoI; The mRNA for the testis library was made

```

from testes and seminal vesicles' hand dissected from 0-3 day old Ore-R males. RNA kindly provided by the lab of Margaret Fuller. Sized fractionated cDNAs were directly ligated into pOTB7. Plasmid cDNA library."

BASE COUNT	208 a	173 c	145 q	124 t
------------	-------	-------	-------	-------

BASE COUNT	208 a	173 c	145 g	124 t	
ORIGIN					
	Query Match	4.1%	Score 35.8;	DB 149;	Length 650;
	Best Local Similarity	65.8%	Pred. No. 6.8;		
	Matches 52;	Conservative 0;	Mismatches 27;	Indels 0;	Gaps 0;
Qy	16	ctggacacatttgcctccaatcattgatgaaggcagcttcgaaggcgccctcttagccctt	75		
Db	188	CTGGACAAGCTGCTCTCCATGATAGCGAAGCAGCAGGAGAGGCCCTCTCAACATA	247		
Qy	76	tccattttccccctcgcg	94		
Db	248	TCCACCGACACTTGGCAG	266		

RESULT	7
AQ851405	
LOCUS	393 bp DNA GSS
DEFINITION	LMAJFVl.lm40bl1.y1 Leishmania major FvI random genomic library Leishmania major genomic clone LMAJFVl.lm40bl1 5', DNA sequence.
ACCESSION	AQ851405
VERSION	AQ851405.1 GI:6056053
KEYWORDS	GSS.
SOURCE	Leishmania major. Leishmania major Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
REFERENCE	1 (bases 1 to 393)
AUTHORS	Akopyants,N.S., Clifton,S.W., Martin,J., Pape,D., Wylie,T., Li,L., Kissinger,J., Roos,D.S., Marra,M., Hillier,L., Chinwalla,A., Blisstein,A., Schmitt,A., Person,B., Theising,B., Ritter,E., Ronko I., Bennett,J., Cole,R., Underwood,K., Cardenas,M., Gibbons,M., Harvey,N., McCann,R., Tsagariswill,R., Williams,T., Jackson,Y., Bowers,Y., Swaller,T., Waterston,R., Wilson,R. and Beverley,S.M. A survey of the Leishmania major Friedlin strain V1 genome by shotgun sequencing: a resource for DNA Mol. Biochem. Parasitol. 113 (2), 337-340 (2001) Other_GSSs: lm40bl1.x1
TITLE	Contact: Akopyants, NS / Beverley, SM
JOURNAL	WashU Leishmania Project
COMMENT	Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA Tel: 314 286 1800 Fax: 314 286 1810

```

Library construction: Natalia S. Akopyants, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
If using this information please cite:
N.S. Akopyants and S.M. Beverley 'A survey of the Leishmania major
Friedlin strain V1 genome by shotgun sequencing' and the Washington
University Genome Sequencing Center For information on obtaining
clone material please contact: Natalia S. Akopyants Ph.D.
(natalia@borcim.wustl.edu) and/or Stephen M. Beverley Ph.D.
(beverley@borcim.wustl.edu)
Seq primer: -40RP from Gibco
Class: shotgun
High quality sequence stop: 357.
Location/Qualifiers
1. .393
/organism="Leishmania major"
/strain="Friedlin strain V1"
/db_xref="taxon:5664"
/clone="LMAJFV1_lm40b11"
/clone_lib="Leishmania major FV1 random genomic library"
/lab_host="TOP10 (Invitrogen)"
/note="Vector: pzero-2 (Invitrogen)"
Genomic DNA was isolated from stationary phase cells. For

```

105 a 97 c 116 g 75 t

BASE COUNT
ORIGIN

this library, DNA was sheared to give a tight size distribution of 1-1.5kb fragments, blunt-ended with T4 DNA polymerase, dephosphorylated with Shrimp Alkaline Phosphatase and ligated into pzero-2 vector's EcoRV site."

	Query Match	4.18;	Score 35.6;	DB 234;	Length 393;
	Best Local Similarity	53.6%;	Pred. No. 6.8;	Mismatches 64;	Indels 0; Gaps 0;
	Matches 74;	Conservative	0;		
Qy	159	ggccaaagcaaccgaacgagggtgaagtcccttgtccaagcagcgcggaataatgggttgct	218		
Dd	185	gtccgttagcgacagtgtgcgagaacctcacccccaagcagcgagtcgttgggtccggct	244		
Qy	219	gcaagcgaaaactaaagcgcaactatctggacgccttgcrgaaatcccgtttaaccatcgc	278		
Dd	245	actagcagtaatccgacagcagcaaacggggagggatcgtgaagacccttggttagccacggc	304		
Qy	279	catcaacgacgattcgcgt	296		
Dd	305	actcaagcgcaatctgat	322		

RESULT	8
LOCUS	AZ135253/c
DEFINITION	AZ135253 660 bp DNA GSS 02-JUN-2000 OSJNB0115C04f CUGI Rice BAC Library (EcoRI) Oryza sativa genomic clone OSJNB0115C04f, DNA sequence.
ACCSSION	AZ135253
VERSION	AZ135253.1 GI:8214549
KEYWORDS	GSS.
SOURCE	Oryza sativa. Oryza sativa
ORGANISM	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae; Oryza. 1 (bases 1 to 660) Wing, R.A. and Dean, R.A. A BAC End Sequencing Framework to Sequence the Rice Genome Unpublished (1998) Contact: Wing RA
REFERENCE	
AUTHORS	
TITLE	
JOURNAL	
COMMENT	

Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwine@clemson.edu
Seq primer: GTAAACGACGGCCAGTG
Class: BAC ends
High quality sequence start: 39
High quality sequence stop: 590.

FEATURES
source 1..660
Location/Qualifiers

```

/organism="Oryza Sativa"
/strain="Japanica"
/cultivar="Nipponbare"
/db_xref="taxon:4530"
/clone="OSJNBb0115C04f"
/clone_lib="CUGI Rice BAC Library (EcoRI)"
/tissue_type="Leaf"
/lab_host="E. coli DH10B"
/notes="Vector: pBACIndigo; Site_1: EcoRI; Site_2: EcoRI;
Rice is the most important food crop in the world. Half of
the world population, especially those inhabiting highly
populated areas of the humid tropics and subtropics, rely
on rice as their primary source of carbohydrate.
Monocotyledonous rice is a diploid plant (2n=24) with a
haploid genome equivalent of 431 Mbp (Arumuganathan and
Earle, 1991). The relatively small genome of rice, three
times larger than that of Arabidopsis, makes it suitable
for genomic studies. In order to facilitate positional

```

cloning, physical mapping and genome sequencing of rice, we have constructed a BAC library from *Oryza sativa*, Nipponbare variety using EcoRI as the cloning enzyme. The library contains 55,296 clones with an average insert size of 121 kb providing approximately 15 haploid genome equivalents. The deep coverage allows the isolation a particular sequence with a probability of 99.9 %. Three high density filters, each containing 18,432 clones (doubly spotted), represent the whole library for colony screening and can be requested from the Clemson University BAC/EST Resource Center (www.genome.clemson.edu).

```

BASE COUNT      121 a   227 c   152 g   156 t      4 others
ORIGIN

Query Match      4.1%; Score 35.6; DB 238; Length 660;
Best Local Similarity 57.0%; Pred. No. 7.8;
Matches 65; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

Qy 529 gggaaactagattgggtgcgattcccgcttaccgttcggtcccaagatgtgtctcaa 588
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 549 GGGGAACAAGGTGTGTTCACCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 490
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 589 gaccgtacactggacggcgctgattgctctgtggggcgacggcggtatcc 642
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 489 GCGCGGACCAAGAAGCCGTCGATCGTCTCCGCGGCGGCGGCGGCGGCGGCGGCGGCGG 436
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 9
AZ127284/c      835 bp      DNA      GSS      02-JUN-2000
LOCUS
DEFINITION      OSJNB0084J06f CUGI Rice BAC Library (EcoRI) Oryza sativa genomic
clone OSJNB0084J06f, DNA sequence.
ACCESSION      AZ127284
VERSION      AZ127284.1 GI:8202809
KEYWORDS      GSS.
SOURCE      Oryza sativa.
ORGANISM      Oryza sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzeae; Oryza.
REFERENCE      1 (bases 1 to 835)
AUTHORS      Wing, R.A. and Dean, R.A.
TITLE      A BAC End Sequencing Framework to Sequence the Rice Genome
JOURNAL      Unpublished (1998)
COMMENT      Contact: Wing RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Seq primer: GTAAACGAGCGGCACGTG
Class: BAC ends
High quality sequence start: 34
High quality sequence stop: 745
Location/Qualifiers
1. .835
/organism="Oryza sativa"
/strain="Japanica"
/cultivar="Nipponbare"
/db_xref="taxon:4530"
/clone="OSJNB0084J06f"
/clone_lib="CUGI Rice BAC Library (EcoRI)"
/tissue_type="Leaf"
/lab_host="E. coli DH10B"
/notes="Vector: pBACIndigo; Site_1: EcoRI; Site_2: EcoRI;
Rice is the most important food crop in the world. Half of
the world population, especially those inhabiting highly
populated areas of the humid tropics and subtropics, rely
on rice as their primary source of carbohydrate.
Monocotyledonous rice is a diploid plant (2n=24) with a
haploid genome equivalent of 431 Mbp (Arumuganathan and

```

Earle, 1991). The relatively small genome of rice, three times larger than that of Arabidopsis, makes it suitable for genomic studies. In order to facilitate positional cloning, physical mapping and genome sequencing of rice, we have constructed a BAC library from *Oryza sativa*, Nipponbare variety using EcoRI as the cloning enzyme. The library contains 55,296 clones with an average insert size of 121 kb providing approximately 15 haploid genome equivalents. The deep coverage allows the isolation a particular sequence with a probability of 99.9 %. Three high density filters, each containing 18,432 clones (doubly spotted), represent the whole library for colony screening and can be requested from the Clemson University BAC/EST Resource Center (www.genome.clemson.edu).

```

BASE COUNT      133 a   287 c   202 g   210 t      3 others
ORIGIN

Query Match      4.1%; Score 35.6; DB 238; Length 835;
Best Local Similarity 57.0%; Pred. No. 8.3;
Matches 65; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

Qy 529 gggaaactagattgggtgcgattcccgcttaccgttcggtcccaagatgtgtctcaa 588
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 548 GGGGAACAAGGTGTGTTCACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 489
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 589 gaccgtacactggacggcgctgattgctctgtggggcgacggcggtatcc 642
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 488 GCGCGGACCAAGAAGCCGTCGATCGTCTCCGCGGCGGCGGCGGCGGCGGCGGCGGCGG 435
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 10
BE605012/c      592 bp      mRNA      EST      21-AUG-2000
LOCUS
DEFINITION      WHE1713-1716_A23_A23ZS Wheat heat stressed spike cDNA library
Triticum aestivum CDNA clone WHE1713-1716_A23_A23, mRNA sequence.
ACCESSION      BE605012
VERSION      BE605012.1 GI:9862282
KEYWORDS      EST.
SOURCE      bread wheat.
ORGANISM      Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Triticeae; Triticum.
REFERENCE      1 (bases 1 to 592)
AUTHORS      Anderson, O.D., Chao, S., Choi, D.W., Close, T.J., Fenton, R.D., Han
, P.S., Hsia, C.C., Kang, Y., Lazo, G.R., Malatrasi, M., Miller, R.,
Nguyen, H.T., Rausch, C.J., Seaton, C.L., Tong, J.C. and Zhang, D.
TITLE      The structure and function of the expressed portion of the wheat
genomes - Heat stressed spike cDNA library
JOURNAL      Unpublished (2000)
COMMENT      Contact: Olin Anderson
US Department of Agriculture, Agriculture Research Service, Pacific
West Area, Western Regional Research Center
800 Buchanan Street, Albany, CA 94710, USA
Tel: 5105959773
Fax: 5105959818
Email: candern@pw.usda.gov
Sequence have been trimmed to remove vector sequence and low
quality sequence with phred score less than 20
Seq primer: Stratagene SK primer.
Location/Qualifiers
1. .592
/organism="Triticum aestivum"
/cultivar="Chinese Spring"
/db_xref="taxon:4565"
/clone="WHE1713-1716_A23_A23"
/clone_lib="Wheat heat stressed spike cDNA library"
/tissue_type="Whole spike"
/dev_stage="Spikes at 5, 10, 15 and 20 days after
anthesis"
/lab_host="E. coli SOLR"
/notes="Vector: Lambda Uni-ZAP XR, excised phagemid;

```

using E. coli host strain XL1 Blue MRF'. Inserts were then subcloned by mass excision using EXAssist helper phage for conversion into phagemid vector pBK-CMV in E. coli host strain XL0LR. Sequenced using T3 primer: 5' ATT AAC CCT CAC TAA AGG GA 3'."

BASE COUNT	93 a	173 c	173 g	96 t	2 others
ORIGIN					
Query Match	4.0%; Score 35; DB 137; Length 537;				
Best Local Similarity	52.1%; Pred. No. 11;				
Matches 100; Conservative	0; Mismatches 91; Indels 1; Gaps 1;				

Db	434	CTTTGATGGCGCGCCGNCGGTCCACGACGGGGATCTCGACACGAAAGATGCAGGCCGT	375
QY	595	gacctgaacggcgctcgatggtcctgtggtggcgacggcgctatccattgtcccgctcg	654
Db	374	GACCGAGGCGCCGCCCT-GGTGACGAGGGCGATGCGGGCTCGCAGTCCGCCCGTGG	316
QY	655	gcggaaggttttgtaggcaattccgcgagggccttggttggggacttcttcccgaaacc	714
Db	315	CCAGTAGGTGTCACGAGCAGCAGCGGGTCGCCCTGTTTGACGGGGTACGCCCTGGATGC	256
QY	715	caagctgctccc	726
Db	255	TCAGGTGGTCGC	244

CNS000E83	749 bp	DNA	GSS	04-JUN-1999
LOCUS				
DEFINITION	Drosophila melanogaster genome survey sequence T7 end of BAC # BACR28N12 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.			
ACCESSION	AL069062			
VERSION	AL069062.1	GI:4949022		
KEYWORDS	GSS.			
SOURCE	fruit fly.			
ORGANISM	Drosophila melanogaster			
	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.			
REFERENCE	1 (bases 1 to 749)			
AUTHORS	Genoscope.			
TITLE	Direct Submission.			
JOURNAL	Submitted (02-JUN-1999) Genoscope - Centre National de Sequençage : Bp 191 91006 EVRY cedex - FRANCE (E-mail : segref@genoscope.cns.fr - Web : www.genoscope.cns.fr)			

FEATURES	SOURCE	BASE COUNT
Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Oosagawa and Aaron Mammoss in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm .	Location/Qualifiers	
	1. - 749	
	/organism="Drosophila melanogaster"	
	/db_xref="taxon:7227"	
	/clone_lib="RPCI-98"	
	/clone="BACR28N12"	
	/note="end : T"	
	78 a 80 c 81 g 96 t 414 others	

KEYWORDS EST.

Search completed: September 14, 2001, 06:59:27
Job time: 12074 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compuqen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 14, 2001, 07:51:13 ; Search time 93.36 Seconds
(without alignments)
1770.229 Million cell updates/sec

Title: US-09-105-117I-1 COPY 1421 2293

Perfect score: 873

Sequence: 1 atgaacccattcaactgga.....tcgagggaattgcggccttag 873

Scoring table: ' IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 324599 seqs, 94655562 residues

Total number of hits satisfying chosen parameters: 649198

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

4
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843

Post-processing: Minimum Match 0%
Maximum Match 100%

Maximum Match 100%
Listing first 45 summaries

SECRET

Database : Issued_Patents_NA:*

```
1: /cgnl_7/ptodata/1/ina/5A_COMB.seq:*
```

```
2: /cgnl//pdata/T/1na/3v_comb.seq:4
3: /cgnl//pdata/1/1na/6v_comb.seq:4
```

```
3: /cgn1_7/ptodata/1/ina/6A_COMB.seq:*
4: /cgn1_7/ptodata/1/ina/6B_COMB.seq:*
```

5: /cgn1_7/ptodata/1/ina/PCTUS_COMB.se

6: /cgn1_7/ptodata/1/ina/backfiles1.se

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %		Length	DB	ID	Description
		Match					
1	122.6	14.0	15239	1	US-08-390-878-17	Sequence 17, Appl	
C 2	37	4.2	2058	1	US-08-358-117-1	Sequence 1, Appli	
C 3	37	4.2	2058	3	US-08-470-588-1	Sequence 1, Appli	
C 4	35	4.0	38506	3	US-09-320-878-19	Sequence 19, Appl	
C 5	33	3.8	379	4	US-08-205-697A-27	Sequence 27, Appl	
C 6	33	3.8	379	5	PCT-US95-02576-27	Sequence 27, Appl	
C 7	33	3.8	1491	2	US-08-147-773-1	Sequence 1, Appli	
C 8	33	3.8	1491	2	US-08-456-104-5	Sequence 5, Appli	
C 9	33	3.8	1491	2	US-08-101-624-22	Sequence 22, Appl	
C 10	33	3.8	1491	2	US-08-751-767A-5	Sequence 5, Appli	
C 11	33	3.8	1491	3	US-08-153-263-1	Sequence 5, Appli	
C 12	33	3.8	1491	3	US-08-479-744A-28	Sequence 28, Appl	
C 13	33	3.8	1491	3	US-08-280-757B-28	Sequence 28, Appl	
C 14	33	3.8	1491	3	US-09-159-135-1	Sequence 1, Appli	
C 15	33	3.8	1491	4	US-08-205-697A-18	Sequence 18, Appl	
C 16	33	3.8	1491	5	PCT-US95-02576-18	Sequence 18, Appl	
17	33	3.8	2656	2	US-08-685-625A-5	Sequence 5, Appli	
C 18	32.4	3.7	1539	3	US-08-582-740-69	Sequence 69, Appl	
C 19	31	3.6	1599	3	US-08-288-405A-9	Sequence 9, Appli	
C 20	31	3.6	4700	4	US-09-150-460B-9	Sequence 9, Appli	
C 21	30.8	3.5	1624	3	US-08-582-740-67	Sequence 67, Appl	
C 22	30.8	3.5	8501	3	US-08-793-900-1	Sequence 1, Appli	
C 23	30.6	3.5	1299	4	US-09-385-028-21	Sequence 21, Appl	
C 24	30.6	3.5	11604	4	US-09-385-028-13	Sequence 13, Appl	
C 25	30.6	3.5	15079	4	US-09-385-028-1	Sequence 1, Appli	
C 26	30.2	3.5	1107	1	US-08-299-849B-23	Sequence 23, Appl	
C 27	30.2	3.5	1107	2	US-08-142-368A-23	Sequence 23, Appl	

ALIGNMENTS

RESULT 1

```

US-08-390-878-17
; Sequence 17, Application US/08390878
; Patent No. 5700683
; GENERAL INFORMATION:
; APPLICANT: Stover, Charles K.
; APPLICANT: Mahaitas, Gregory G.
; TITLE OF INVENTION: VIRULENCE-ATTENUATING GENETIC DELETIONS
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend Khourie and Crew
; STREET: One Market Plaza, Steuart Street Tower, 20th
; STREET: Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:

```

APPLICATION NUMBER: US/08/390,878
FILING DATE: 17-FEB-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Hunter, Tom
REGISTRATION NUMBER: 38,498
REFERENCE/DOCKET NUMBER: 15371A-17
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/543/9600
TELEFAX: 415/543/5043
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 15239 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-390-878-17

Query Match	14.0%;	Score 122.6;	DB 1;	Length 15239;
Best Local Similarity	49.9%;	Pred. No. 8.3e-28;		

Qy 13 caactggacactttgctctcaatcattgatgaagcagcttcgaagcgccctccttagcc 72
 ||||| | ||| - | ||| ||||| ||| |
 Db 5253 CAGCTGGCCCGAATTGGCTGCCGTGGTGCAGAACCTGGCAGCTTCGATGGCGCCGGGAGCGC 5312

```

QY 73 cttccattccccctcgcgcgtgagtcagcgcgttaaaagctctcagcatcacgtgggt 132
Db 5313 CTACATGTCACCCCTCGCGCTGTCAGTCAGGCATCAATGCTGTGGAGCAGCAGGTCGCG 5372
QY 133 cagtggtggtatcgcgaccccaacgcgcgaagcaaacgagcgcgcgtgagtcctgtg 192
Db 5373 CAGGTGCTGCTGTCAGGGAAGCCATGTCGGCGCAGCAGCGATATCCGCTGTTG 5432
QY 193 caagcagcgcgcgaagtggtgctgcaagcagcaaaactaaagcgcgaactatcttg -ac 250
Db 5433 CGGTGCGCGCGCAACACAGCGTGTCTGAGTCCGAGCGCTGCTGAAATGGTGGCAAC 5492
QY 251 gcttgctgaa----atccgcttaacccatcgccatcaacgagatcgatccacatgg 306
Db 5493 GCGTCTGAACGACGCGGGATCACCATTGGGTAAACCGCGATTCATCGGGACATGG 5552
QY 307 ttctcccggtgttcaacgagtagcttcttgggtggagcaacgctcagctggtgctg 366
Db 5553 TTTTCGGCGGTGTTGACG-----GTCGCGCGAGCTGCTGCTGACGTTGCGATC 5603
QY 367 gaagatgaagcagcacattatcttctgctgcgctgagatgttttagagcggttaacc 426
Db 5604 GAGGACCGAGACCATTCGCGCGGCTGTCACGGAGGGTGTGCGGATGGCGCGGTGACG 5663
QY 427 cgtgaagctaatcccggtgcggtgagtagtagtagaacttggaaccatcgccacttg 486
Db 5664 ACCGACCGGAACCCGTCGCGGCTGCCGGGTGCACCGCTGGGTGAAATGCGCTACCTA 5723
QY 487 gcaatgaacccccctcattgctggtgagtgctactatggttgatggaaactagatgggt 546
Db 5724 CCAGTGGCGCAGCAGGCAATTCGTCAGCGCCATCTATCCGACGGTTACTGCCGCGCG 5783
QY 547 gcaatgcccgttaccgcttgcgttcccaagatggtcttcaagacgctgacgtgacgg 606
Db 5784 GCGGTAAAGCTCCGTCACCTGGCGTGGGAATCGTGAGATGGGCTGAGGACATGTTGGTG 5843
QY 607 cg---cgtcgatggtctgtggtggtgagcgcgtatccattgtccgttcggtcggaagt 663
Db 5844 CGTAAGGCTTTTCGTCGCGCCATCACCACAGCAGCAGCTTTGTCGCGACACAGAGGC 5903
QY 664 ttggtgaggaattgcgcgcgcgttgggtggtggtggtggtggtggtggtggtggtggt 723
Db 5904 TACCGCGCGCAGCGCGCGCGGTGGATGGGCTGTCGCGGATGTCGCGGAGAGTGGCAGCA 5963
QY 724 cccatgctaaagcaggaagtagtctcctcgtgagatgagatgagatgagatgagatg 783
Db 5964 TCTCCGCTTGGCGATGGATCGTTCGTCGAGGCTGCGGATACACCTCGACGTCCTCTC 6023
QY 784 tattggaacgagtcgctggaatagatctctagctagactacagacgcgcgtcggt 843
Db 6024 TATTGCAATGCTGGAACCTGGACATCGCATCATCGCGCGAATTACCGACAGCGGTGAG 6083
QY 844 gatgaagaatcg 856
Db 6084 GCGGCGGAAGCG 6096

```

RESULT 2

US-08-358-117-1/c
; Sequence 1, Application US/08358117
; Patent No. 5608147

GENERAL INFORMATION:

; APPLICANT: Kaphammer, Bryan J.
; TITLE OF INVENTION: tda Gene Selectable Markers in Plants and the
; TITLE OF INVENTION: Use Thereof
; CORRESPONDENCE ADDRESS:
; ADDRESS: Sterne, Kessler, Goldstein & Fox
; STREET: 1100 New York Avenue, Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA

; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/358,117
; FILING DATE: 15-DEC-1994
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Esmond, Robert W.
; REGISTRATION NUMBER: 32,893
; REFERENCE/DOCKET NUMBER: 1405.0030001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2058 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 751..1611
US-08-358-117-1

Query Match 4.2%; Score 37; DB 1; Length 2058;
Best Local Similarity 53.9%; Pred. No. 0.073;
Matches 76; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

```

QY 13 caactggacacatttctcctcaatcattgaagcagcttcgaagcgcctccttagcc 72
Db 530 CAGCTTCGCTATTTCCTGCTGCGGAGGAGGCAACGTCGTCGCCCGCGCGCGG 471
QY 73 cttccattccccctcgcgcgtgagtcagcgcgttaaaagctctcagcatcacgtgggt 132
Db 470 CTGCAATATTCACGCCCGCGGTACGCGCAGACAGATTACGCGCTCGACACGATCTGGGC 411
QY 133 cagatgttggtatcgccacc 153
Db 410 GTGTTGTTGTTGAGCGCAGC 390

```

RESULT 3

US-08-470-588-1/c
; Sequence 1, Application US/08470588
; Patent No. 6100446

GENERAL INFORMATION:

; APPLICANT: STREBER, WOLFGANG R.
; APPLICANT: TIMMIS, KENNETH H.
; TITLE OF INVENTION: MICROORGANISMS AND PLASMIDS FOR
; TITLE OF INVENTION: 2,4-DICHLOROPHENOXACETIC ACID (2,4-D) MONOOXYGENASE
; TITLE OF INVENTION: FORMATION AND PROCESS FOR THE PRODUCTION OF THESE PLASMIDS
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESS: FROMMER LAWRENCE & HAUG LLP
; STREET: 745 FIFTH AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10151
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/470,588

Query Match 4.2%; Score 37; DB 3; Length 2058;
Best Local Similarity 53.9%; Pred. No. 0.073;
Matches 76; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

QY 13 caactggacacttgcctcaatcattgatgaaggagcgttcgaaggcgcctccttagcc 72
DB 530 CAGCTTCGCTATTTCGTGTCTGCGCGGAGGAGGCAACGTGCGTCCGCCGCGCGCGG 471
QY 73 ctttcattccccctcgcgcgttgagtcagcgcgtttaaagctctcagacatcacgttggt 132
DB 470 CTGCATATTTCCAGCCCCCGGTTCACCGGACAGATTTCACGGCTCGAAGACATCTGGC 411
QY 133 c9agtggtggtatcgccacc 153
DB 410 GTGTGTGTGTTCGAGCGCAGC 390

RESULT 4
US-09-320-878-19/c
; Sequence 19, Application US/09320878A
; Patent No. 6117659
; GENERAL INFORMATION:
; APPLICANT: ASHLEY, Gary
; APPLICANT: BETLACH, Melanie C.
; APPLICANT: BETLACH, Mary C.
; APPLICANT: MCDANIEL, Robert
; APPLICANT: TANG, Li
; FILE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
; CURRENT APPLICATION NUMBER: US/09/320,878A
; FILING DATE: 1999-05-27
; EARLIER APPLICATION NUMBER: CIP OF 09/141,908
; EARLIER FILING DATE: 1998-08-28
; EARLIER APPLICATION NUMBER: CIP OF 09/073,538
; EARLIER FILING DATE: 1998-05-06
; EARLIER APPLICATION NUMBER: CIP OF 08/846,247
; EARLIER FILING DATE: 1997-04-30
; EARLIER APPLICATION NUMBER: 60/119,139
; EARLIER FILING DATE: 1999-02-08
; EARLIER APPLICATION NUMBER: 60/100,880
; EARLIER FILING DATE: 1998-09-22
; EARLIER APPLICATION NUMBER: 60/087,080
; EARLIER FILING DATE: 1998-05-28
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 19
; LENGTH: 38506
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-320-878-19

Query Match 3.8%; Score 33; DB 4; Length 379;
Best Local Similarity 46.4%; Pred. No. 0.6;
Matches 108; Conservative 0; Mismatches 125; Indels 0; Gaps 0;

QY 406 gatgttttagagcggtaaacccgtgaagctaaatcccggtggcggtatgtgaagtagtagaa 465
DB 28123 GAGTCTGTCGAGGAGCGAGACCGCGCGGTCAAGCCGTCGCCGTCGTCAGTGGGTGAG 28064
QY 466 cttggaaccattgcccacttgccattggccattgcaacccccctcattggggatgctacatggtt 525
DB 28063 CCGGGCGCGGAGGCGCTCAGGGTCTGCTCGCCCGCGGTTCAGAGTACGTACGTCCGC 28004
QY 526 gatggaaactagattggcgatcccgctcttaccgcttcggtcccaaaagatgtgctt 585
DB 28003 TCCGCGGCGCTTACGCGCGCCACAGCTCGGCTTCGGCTCGCTCCCGAGCGGACGAT 27944
QY 586 caagaccgtgacctggagcgcgctcgatggtctctgtggcgagcgcgatccatt 645
DB 27943 CCACCGCGCGACAGGTCGCGCTGGCGGAGCGCGTTCAGCGGCTTCACACTCGACGCGGTA 27884
QY 646 gtcccgtcggcggaaggttttggtag 672
DB 27883 ACGCAGCTGTCGCGCGCGCTGGTGG 27857

RESULT 5
US-08-205-697A-27/c
; Sequence 27, Application US/08205697A
; Patent No. 6218510
; GENERAL INFORMATION:
; APPLICANT: Sharpe, Arlene H.
; APPLICANT: Borriello, Francescopaulo
; APPLICANT: Freeman, Gordon J.
; APPLICANT: Nadler, Lee M.
; TITLE OF INVENTION: No. 6218510el Forms of T Cell Costimulatory Molecules
; TITLE OF INVENTION: and Uses Therefor
; NUMBER OF SEQUENCES: 61
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 State Street, suite 510
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/205,697A
; FILING DATE: 02-Mar-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Mandragouras, Amy E.
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: BWI-120
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)227-5941
; INFORMATION FOR SEQ ID NO: 27:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 379 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..69
US-08-205-697A-27

Query Match 3.8%; Score 33; DB 4; Length 379;
Best Local Similarity 46.4%; Pred. No. 0.6;
Matches 108; Conservative 0; Mismatches 125; Indels 0; Gaps 0;

Db	270	GAAAGAGAGAGGAGGAGCCACTTCTGTAACCTGTCAGCTTAATCTGGAAGAGGTTT	271	TTA
Qy	301	acatggttctcccgctggttcaacgagggtagctctcttgggggtggagcaacgctcacgctg	302	
Db	210	AGAAGCTGAGGTAGTGGGTAAAAACAGCTTAAATTTGTTAAGGGAAGAAATGCCCTCATGATC	211	151
Qy	361	cgcttggaaatgaagcgcacacatctatccctgtcgcgcgtggagatgttt	362	413
Db	150	CCACATCATCATATCCCAAGAGAGATGACGGAGGCTACCTTCAGATCTTTT	151	98

```

RESULT 7
US-08-147-772-1/c
; Sequence 1, Application US/08147772
; Patent No. 5858776
; GENERAL INFORMATION:
; APPLICANT: Ostrand-Rosenberg, Suzanne
; APPLICANT: Baskar, Sivasubramanian
; APPLICANT: Glincher, Laurie H.
; APPLICANT: Freeman, Gordon J.
; APPLICANT: Nadler, Lee M.
; TITLE OF INVENTION: Tumor Cells With Increased Immunogenicity
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 State Street, Suite 510
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/147,772

```

CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Mandragouras, Amy E.
REGISTRATION NUMBER: 36,207
REFERENCE/DOCKET NUMBER: RPI-003
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 227-7400
TELEFAX: (617) 227-5941
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1491 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
HYPOTHETICAL: no
ANTI-SENSE: no
ORIGINAL SOURCE:
ORGANISM: Homo sapien
TISSUE TYPE: lymphoid
CELL TYPE: B cell
CELL LINE: Raji
IMMEDIATE SOURCE:
LIBRARY: cDNA in pCDM8 vector
CLONE: B7, Raji clone #13
POSITION IN GENOME:
CHROMOSOME/SEGMENT: 3
FEATURE:
NAME/KEY: Open reading frame (translated region)
LOCATION: 318 to 1181 bp
IDENTIFICATION METHOD: similarity to other patterns

APPLICATION NUMBER: 19-AUG-1993
 ATTORNEY/AGENT INFORMATION:
 NAME: Mandragoras, Amy E.
 REGISTRATION NUMBER: 36,207
 REFERENCE/DOCKET NUMBER: RPI-008
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (617) 227-7400
 TELEFAX: (617) 227-5941
 INFORMATION FOR SEQ ID NO: 5:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1491 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: cDNA to mRNA
 ORIGINAL SOURCE:
 ORGANISM: Homo sapien
 TISSUE TYPE: lymphoid
 CELL TYPE: B cell
 CELL LINE: Raji
 US-08-456-104-5

	Query Match	3.88;	Score 33;	DB 2;	Length 1491;
	Best Local Similarity	46.4%;	Pred. No. 1.1;		
	Matches 108;	Conservative	0;	Mismatches 125;	Indels 0;
	Gaps				
QY	181	gaagtcccttg	tgaacgacgcgcg	gaaaaatg	gtgttctg
Db	1442	GATGGCAGAA	TGGAAACATG	GCAAAAGAGG	TTACATTAA
QY	241	ctatctggacgc	cttgc	tgaatccgc	cttaaccatcg
Db	1382	GAAAGGAGAG	GATGCCAC	CTCTTAACTG	TTCACGCTTA
QY	301	acatggttcc	ctccgtgtt	ctcaacagc	aggttagctt
Db	1322	AGAAGGTG	AGGTAGTGG	TGGTAAAAC	AGCGTTAAAT
QY	361	cgcttgaag	atgaacgcgc	acacattc	ctctgctgcgc
Db	1262	CCACAGT	ATCATATCC	CAGAAGAGAT	GACGGAGGCT

```

RESULT          9
US-08-101-624-22/c
; Sequence 22, Application US/08101624
; Patent No. 5942607
;
; GENERAL INFORMATION:
; APPLICANT: Freeman, Gordon J.
; APPLICANT: Nadler, Lee M.
; APPLICANT: Gray, Gary S.
; TITLE OF INVENTION: No. 5942607el CTLA4/CD28 Ligands and
; TITLE OF INVENTION: Uses Therefor
; NUMBER OF SEQUENCES: 25
;
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 State Street, Suite 510
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/101,624
; FILING DATE: 26-JUL-1993
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:

```

Query Match 3.8%; Score 33; DB 2; Length 1491;
Best Local Similarity 46.4%; Pred. No. 1.1;
Matches 108; Conservative 0; Mismatches 125; Indels 0; Gaps 0;

	Query Match	3.8%;	Score 33;	DB 2;	Length 1491;
	Best Local Similarity	46.4%;	Pred. No. 1.1;		
	Matches 108;	Conservative	0;	Mismatches 125;	Indels 0; Gaps
Qy	181	gaagtccttgtcaacgacgcgcggaaaaatgggttgctgctgaacgcagaactaaagcgcaa	240		
Db	1442	GATGGCAGATTGGAACATGCGCAAGACAGCTTACATTAGCAAAATGCATATGGGA	1383		
Qy	241	ctatctggagcccttgtgtaattcccgtttaaccatgcgcataccaacgcagattcgcctatccc	300		
Db	1382	GAAGAAGAGAGGATGCCAGCCATTCTGTAACTGTTTCAGCTTAATCTGAAGAGGTTTTTA	1323		
Qy	301	acatgatcttccctccgtgtttcaacgaagttagctcttctgggggtggagcaacctcacgcctg	360		
Db	1322	AGAAGTGAGGTAGTGGGTAAACACACTTAAATTTGTTAAGGACAATGCCTCATGATC	1263		
Qy	361	cgcttggaaagatgaacgcacacatatctctgtcgcgcgctggagatgtttt	413		
Db	1262	CCCACATCATCTATGCCAGAAGAGATCACGGAGCTACCTTCAGATCTTTT	1210		

RESULT 11
US-08-153-262-1/c
; Sequence 1, Application US/08153262
; Patent No. 6071716
; GENERAL INFORMATION:

```

; APPLICANT: FREEMAN, GORDON J.
; APPLICANT: FREEDMAN, ARNOLD S.
; APPLICANT: NADLER, LEE M.
; TITLE OF INVENTION: DNA Encoding B7, A New Member
; TITLE OF INVENTION: Of The IgG Superfamily With Unique Expression On
; TITLE OF INVENTION: Activated And Neoplastic B Cells.
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESS: The Dana-Farber Cancer Institute
; STREET: 44 Binney Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02115
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 720kb storage
; COMPUTER: IBM Personal System 2; Model 30
; OPERATING SYSTEM: MS/DOS
; SOFTWARE: WordPerfect 5.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/153,262
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/751,306
; FILING DATE: 28-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: HART, JULIA D.
; REGISTRATION NUMBER: 33132
; REFERENCE/DOCKET NUMBER: DFCI-116.1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (203) 255-8900
; TELEFAX: (203) 259-2846
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1491 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cdna to mRNA
; HYPOTHETICAL: no
; ANTI-SENSE: no
; ORIGINAL SOURCE:
; ORGANISM: Homo sapien
; TISSUE TYPE: lymphoid
; CELL TYPE: B cell
; CELL LINE: Raji
; IMMEDIATE SOURCE:
; LIBRARY: cdna in pCDM8 vector
; CLONE: B7, Raji clone #13
; POSITION IN GENOME:
; CHROMOSOME/SEGMENT: 3
; FEATURE:
; NAME/KEY: Open reading frame (translated region)
; LOCATION: 318 to 1181 bp
; IDENTIFICATION METHOD: similarity to other pattern
; FEATURE:
; NAME/KEY: Alternate polyadenylation signal
; LOCATION: 1474 to 1479 bp
; IDENTIFICATION METHOD: similarity to other pattern
; PUBLICATION INFORMATION:
; AUTHORS: FREEDMAN, GORDON J.
; AUTHORS: FREEDMAN, ARNOLD S.
; AUTHORS: SEGIL, JEFFREY M.
; AUTHORS: LEE, GRACE
; AUTHORS: WHITMAN, JAMES F.
; AUTHORS: NADLER, LEE M.
; TITLE: B7, A New Member Of The Ig Superfamily With
; TITLE: Unique Expression On Activated And Neoplastic B Cells
; JOURNAL: The Journal of Immunology
; VOLUME: 143
; ISSUE: 8
; PAGES: 2714-2722

```

```

; DATE: 15-OCT-1989
; RELEVANT RESIDUES IN SEQ ID NO: 1: FROM -1 TO 1491
; US-08-153-262-1
;
; Query Match 3.88; Score 33; DB 3; Length 1491;
; Best Local Similarity 46.48; Pred. No. 1.1;
; Matches 108; Conservative 0; Mismatches 125; Indels 0; Gaps 0;
;
; Qy 181 gaagtccttgcagcagcgagaaatggtgtgctgcaagcagaaactaaagcgcaa 240
; | | | | | | | | | | | | | | | | | | | | | | | | | | | |
; Db 1442 GATGGCAGATGGAAACATGCGAAAGAGAGGTTACATTAAGCAATTCATATGGGA 1383
; | | | | | | | | | | | | | | | | | | | | | | | | | | | |
; Qy 241 ctatctgagcgcttgcgtgaaatcccggttaaccatcccatcaacgcagatcgctatcc 300
; | | | | | | | | | | | | | | | | | | | | | | | | | | | |
; Db 1382 GAAAGGAGAGGATGCCAGGCATCTTGTAACTGTTACGCTTAATCTGAAGAGGTTT 1323
; | | | | | | | | | | | | | | | | | | | | | | | | | | | |
; Qy 301 acatggttctcccggtgtcaacgagagtagcttcttgggtgagcaacgcacgctg 360
; | | | | | | | | | | | | | | | | | | | | | | | | | | | |
; Db 1322 AGAAGGTGAGGTAGTGGGTAAACACAGCTTAATTTGTTAAGGGAAGATGCTCATGATC 1263
; | | | | | | | | | | | | | | | | | | | | | | | | | | | |
; Qy 361 cgcttgaagatgaagcgcacacattatccttgcgcgcgtggagatgtttt 413
; | | | | | | | | | | | | | | | | | | | | | | | | | | | |
; Db 1362 CCCACGATCCATGTATCCAGAGAGATGACGGAGGCTACCTTCAGATCTTTT 1210
; | | | | | | | | | | | | | | | | | | | | | | | | | | | |
;
; RESULT 12
; US-08-479-744A-28/c
; Sequence 28, Application US/08479744A
; Patent No. 6084067
; GENERAL INFORMATION:
; APPLICANT: Freeman, Gordon J.
; APPLICANT: Nadler, Lee M.
; APPLICANT: Gray, Gary S.
; TITLE OF INVENTION: No. 6084067el CTLA4/CD28 Ligands and
; TITLE OF INVENTION: Uses Therefor
; NUMBER OF SEQUENCES: 55
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD, LLP
; STREET: 60 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/479,744A
; FILING DATE: June 7, 1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/280,757
; FILING DATE: 26-JUL-1994
; APPLICATION NUMBER: 08/109,393
; FILING DATE: 28-AUG-1993
; APPLICATION NUMBER: 08/101,624
; FILING DATE: 26-JULY-1993
; APPLICATION NUMBER: 08/147,773
; FILING DATE: 3-NOV-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Mandragouras, Amy E.
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: RPI-004CP3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 227-5941
; INFORMATION FOR SEQ ID NO: 28:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1491 base pairs
; TYPE: nucleic acid

```

NUMBER OF SEQUENCES: 53
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD
STREET: 60 State Street, Suite 510
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/280,757B
FILING DATE: 26-JUL-1994
CLASSIFICATION: 435

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/101,624
FILING DATE: 26-JULY-1993
APPLICATION NUMBER: 08/109,393
FILING DATE: 19-AUG-1993
APPLICATION NUMBER: 08/147,773
FILING DATE: 3-NOV-1993

ATTORNEY/AGENT INFORMATION:
NAME: Mandragouras, Amy E.
REGISTRATION NUMBER: 36,207
REFERENCE/DOCKET NUMBER: RPI-004CP2

TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 227-7400
TELEFAX: (617) 227-5941
INFORMATION FOR SEQ ID NO: 28:
SEQUENCE CHARACTERISTICS:
LENGTH: 1491 base pairs
TYPE: nucleic acid
STRANDEDNESS: double

TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
HYPOTHETICAL: no
ANTI-SENSE: no

ORIGINAL SOURCE:
ORGANISM: Homo sapien
TISSUE TYPE: lymphoid
CELL TYPE: B cell
CELL LINE: Raji

IMMEDIATE SOURCE:
LIBRARY: cDNA in pCDM8 vector
CLONE: B7, Raji clone #13
POSITION IN GENOME:
CHROMOSOME/SEGMENT: 3

FEATURE:
NAME/KEY: Open reading frame (translated region)
LOCATION: 318 to 1181 bp
IDENTIFICATION METHOD: similarity to other pattern

FEATURE:
NAME/KEY: Alternate polyadenylation signal
LOCATION: 1474 to 1479 bp
IDENTIFICATION METHOD: similarity to other pattern

PUBLICATION INFORMATION:
AUTHORS: FREEDMAN, GORDON J.
AUTHORS: FREEDMAN, ARNOLD S.
AUTHORS: SEGIL, JEFFREY M.
AUTHORS: LEE, GRACE
AUTHORS: WHITMAN, JAMES F.
AUTHORS: NADLER, LEE M.
TITLE: B7, A New Member Of The Ig Superfamily With
TITLE: Unique Expression On Activated And Neoplastic B Cells
JOURNAL: The Journal of Immunology
VOLUME: 143
ISSUE: 8

PAGES: 2714-2722
DATE: 15-OCT-1989
RELEVANT RESIDUES IN SEQ ID NO: 28: FROM 1 TO 1491

поступил в 1990 году, полагая, что он не имеет отношения к делу.

STATE: Massac
COUNTRY: USA

STATE: Massachusetts
COUNTRY: USA

```

; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/205,697A
; FILING DATE: 02-Mar-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Mandragouras, Amy E.
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: BWI-120
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)227-5941
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1491 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 318..1181
; US-08-205-697A-18

```

```

Query Match      3.8%; Score 33; DB 4; Length 1491;
Best Local Similarity 46.4%; Pred. No. 1.1;
Matches 108; Conservative 0; Mismatches 125; Indels 0; Gaps 0;

QY 181 gaagtcctgtgcaagcagcgagaaatgggtgtgtgtgcaagcagaaactaaagcgcaa 240
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1442 GATGGCAGATGGAACATGGCAAAAGAGAGGTTACATTAGCAAAATTCATATCGGGA 1383
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 241 ctatctggagccttgcgtgaaatcccggttaaccatcgccatcaacgcagattcgctatcc 300
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1382 GAAAGGAGAGGGATGCCAGCATCTTTGTAACGTGTCAGCTTAATCTGAAAGAGGTTTTTA 1323
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 301 acatgtttctccctgctgttcaacagaggtagcttcttggggggagcaacgctcagctg 360
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1322 AGAAGGTGAGGTAGTGGGTAACACAGCTTAAATTTCTTAAGGGAAGAAATGCCCTCATGATC 1263
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 361 cgcttgaagatgaagcgacacattatctgtgcgcgcgtggagatgtttt 413
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1262 CCCACGATCCATGATATCCAGAGAGATGACGGAGGCTACCTTCAGATCTTTT 1210
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

Search completed: September 14, 2001, 07:51:47
Job time: 12421 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 14, 2001, 03:41:38 ; Search time 2960.24 Seconds
(without alignments)
3715.094 Million cell updates/sec

Title: US-09-105-117i-1-copy_1016_1726
Perfect score: 711
Sequence: 1 gtgtcatcgggcagcggttgg.....cagattcgctatccacatgg 711

Scoring table: IDENTITY_NUC
Gapop 10_0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues
Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:
1: gb_ba1:
2: gb_ba2:
3: gb_ba3:
4: gb_in1:
5: gb_in2:
6: gb_in3:
7: gb_om:
8: gb_ov:
9: gb_pat1:
10: gb_pat2:
11: gb_ph:
12: gb_pl1:
13: gb_pl2:
14: gb_pl3:
15: gb_pl4:
16: em_ba1:
17: em_ba2:
18: em_fun:
19: em_htgo_hum:
20: em_htgo_inv:
21: em_htgo_rod:
22: em_htg_hum1:
23: em_htg_hum2:
24: em_htg_hum3:
25: em_htg_hum4:
26: em_htg_hum5:
27: em_htg_hum6:
28: em_htg_hum7:
29: em_htg_hum8:
30: em_htg_inv1:
31: em_htg_inv2:
32: em_htg_other:
33: em_htg_rod:
34: em_hum1:
35: em_hum2:
36: em_hum3:
37: em_hum4:
38: em_hum5:
39: em_hum6:
40: em_hum7:
41: em_in:
42: em_om:
43: em_or:

44: em_ov:
45: em_pat:
46: em_ph:
47: em_pl:
48: em_ro:
49: em_sts:
50: em_sy:
51: em_un:
52: em_vi:
53: gb_sts1:
54: gb_sts2:
55: gb_sts3:
56: gb_sy:
57: gb_un:
58: gb_vil:
59: gb_vil2:
60: gb_vil3:
61: gb_vil4:
62: gb_vil5:
63: gb_vil6:
64: gb_vil7:
65: gb_vil8:
66: gb_vil9:
67: gb_vil10:
68: gb_vil11:
69: gb_vil12:
70: gb_vil13:
71: gb_vil14:
72: gb_vil15:
73: gb_vil16:
74: gb_vil17:
75: gb_vil18:
76: gb_vil19:
77: gb_vil20:
78: gb_vil21:
79: gb_vil22:
80: gb_vil23:
81: gb_vil24:
82: gb_vil25:
83: gb_vil26:
84: gb_vil27:
85: gb_vil28:
86: gb_vil29:
87: gb_vil30:
88: gb_vil31:
89: gb_vil32:
90: gb_vil33:
91: gb_vil34:
92: gb_vil35:
93: gb_vil36:
94: gb_vil37:
95: gb_vil38:
96: gb_vil39:
97: gb_vil40:
98: em_ba3:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
C 1	711	100.0	2374	2	CGLYSEG
C 2	711	100.0	2374	9	A93933
C 3	435	61.2	822	9	AX063771
C 4	406	57.1	993	9	AX063767
C 5	81.2	11.4	33285	3	SC5F8
C 6	71.8	10.1	12070	1	AB011413
C 7	63.2	8.9	14844	3	MBU34849
C 8	63.2	8.9	15239	10	I86263

9	62	8.7	10701	1	AE005522	AE005522 Escherich
10	62	8.7	266658	2	AP002563	AP002563 Escherich
C 11	61.6	8.7	38500	3	MTCY39	Z74025 Mycobacteri
C 12	60.4	8.5	636	9	AX030085	AX030085 Sequence
C 13	60.4	8.5	8029	2	ECFADAPK	X14436 Escherichia
14	60.4	8.5	10362	1	AE000375	AE000375 Escherich
15	60.4	8.5	141744	2	ECU28377	U28377 Escherichia
C 16	58.4	8.2	10542	1	AE004852	AE004852 Pseudomon
C 17	52.8	7.4	1804	2	ECRPIAA	X73026 E.coli gene
C 18	52.8	7.4	1815	2	ECOICIA	M62865 E.coli chro
C 19	52.8	7.4	6055	2	ECSERAICI	X66836 E.coli sera
C 20	52.8	7.4	110000	66	AC020850_0	AC020850 Mus muscu
C 21	51.2	7.2	10367	1	AE005521	AE005521 Escherich
C 22	49.2	6.9	11168	1	AE004560	AE004560 Pseudomon
C 23	46.2	6.5	10572	1	AE004780	AE004780 Pseudomon
C 24	45.4	6.4	4513	10	AX067447	AX067447 Sequence
C 25	44	6.2	37218	3	MTCY2069	Z77162 Mycobacteri
C 26	43.6	6.1	3632	1	AF037440	AF037440 Edwardsie
C 27	43	6.0	14789	1	AE004916	AE004916 Pseudomon
C 28	42.8	6.0	13193	1	AE005456	AE005456 Escherich
C 29	42.8	6.0	26857	2	AP002560	AP002560 Escherich
C 30	42	5.9	4532	3	SYOSBRA	L41665 Synecococc
C 31	41.4	5.8	4033	2	AS065741	U65741 Aeromonas s
C 32	40.6	5.7	10010	1	AE004639	AE004639 Pseudomon
C 33	40	5.6	10239	1	AE004457	AE004457 Pseudomon
C 34	40	5.6	183704	67	AC023024	AC023024 Homo sapi
C 35	39.8	5.6	2878	2	AHAHYGENE	X89469 Aeromonas h
C 36	39	5.5	160938	77	AC090056	AC090056 Oryza sat
C 37	38.8	5.5	1033	1	AB053349	AB053349 Synecococ
C 38	38.8	5.5	4217	2	AY017305	AY017305 Synecococ
C 39	38.6	5.4	1070	3	PSECATRA	M33817 P.putida ca
C 40	37.8	5.3	11015	1	AE004736	AE004736 Pseudomon
C 41	37.6	5.3	5644	1	AF061070	AF061070 Pseudomon
C 42	37.2	5.2	10614	1	AE004351	AE004351 Vibrio ch
C 43	37.2	5.2	187111	88	AC087244	AC087244 Homo sapi
C 44	37	5.2	1592	3	S80112	S80112 Ralstonia e
C 45	37	5.2	1790	3	PJPTFDRII	M98445 Ralstonia e

RESULT 1	CGLYSEG	2374 bp	DNA	18-MAR-2001
LOCUS	C.glutamylase	lysE and lysG genes.		
DEFINITION	C.glutamylase	lysE and lysG genes.		
ACCESSION	X96471			
VERSION	X96471.1	GI:1729753		
KEYWORDS	lysE gene; lysG gene; Lysine export regulator protein; Lysine exporter protein; Lysine governor.			
SOURCE	Corynebacterium glutamicum.			
ORGANISM	Corynebacterium glutamicum			
REFERENCE	1 (bases 1 to 2374)			
AUTHORS	Vrljic, M., Sahn, H., and Eggeling, L.			
TITLE	A new type of transporter with a new type of cellular function: L-lysine export from Corynebacterium glutamicum			
JOURNAL	Mol. Microbiol. 22 (5), 815-826 (1996)			
MEDLINE	97126810			
REFERENCE	2 (bases 1 to 2374)			
AUTHORS	Vrljic, M.M.			
TITLE	Direct Submission			
JOURNAL	Submitted (07-MAR-1996) M.M. Vrljic, Institut fuer Biotechnologie			
FEATURES	1, Forschungszentrum Juelich, Postfach 1913, D-52425 Juelich, FRG			
source	Location/Qualifiers			
	i. .2374			
	/organism="Corynebacterium glutamicum"			
	/strain="R127"			
	/db_xref="taxon:1718"			
gene	complement(82..954)			
	/gene="lysG"			

CDS	complement(82..954)			
	/gene="lysG"			
	/note="lysine governor"			
	/codon_start=1			
	/transl_table=11			
	/product="lysine export regulator protein"			
	/protein_id="CAA65323.1"			
	/db_xref="GI:1729754"			
	/db_xref="SWISS-PROT:P94632"			
	/translation="MNPQLDITLLSIIDEGSGFEGASLALSTSPSAVSORVRALEHHVG			
	RVLSRTQPAKATEAGEVLQAKRMVLLQAEATKAQLSRLAEIPLTIAINADLSUTW			
	FPFVNEVASMGATLTLEDEAHTLSILRRGDVLGATREANFVACVEVGLTMR			
	HLAATPSLRDAYMVDGLDWAAMPVLRFGKDVLDQLDRLDGRVDPGVRRRVSTVPS			
	AEFGCEAIRRGIGWGLLPETQAAPMLKAGEVILLDEIPDTPMYQWRLESRLARL			
	TDAVYDAIEGLRP"			
	1025..1726			
gene	/gene="lysE"			
	1025..1726			
CDS	/gene="lysE"			
	/codon_start=1			
	/transl_table=11			
	/product="lysine exporter protein"			
	/protein_id="CAA65324.2"			
	/db_xref="GI:13397387"			
	/db_xref="SWISS-PROT:P94633"			
	/translation="MEIFITGLLGASLLSTGPONVILVIRKQIGREGLIAVLLVCLLI			
	SDVFLFIATGLVDLLSNAAPITVLDIMRWGGTAYLLWFAVNAKDKAMTKVAPQIIIE			
	ETEPVDPDTPLGGSATVATDRNRVRVSVSDKORVVKPMLMALVTLWLPNAYLDA			
	FVFVGGVAGVGDYGRWTFAPAGAFASLIWFLPLVGFGAALSRPLSSPKVRWVNVV			
	AVVMTALAIKLMVG"			
	complement(1723..2352)			
gene	/gene="orf3"			
CDS	complement(1723..2352)			
	/gene="orf3"			
	/codon_start=1			
	/transl_table=11			
	/protein_id="CAA65325.1"			
	/db_xref="GI:1729756"			
	/db_xref="SPTREMBL:P94634"			
	/translation="MYALRDIVASGKALYVGISSYSGPELTAAEAEEFMAEEGCLLIHQ			
	PSYIINRWVEPGDDGNLLQSAANGLGVIAFSPLAGLLTDKYLDIPEGSSRASQ			
	GKLSSEGMNVNNDIMVRKLDIAQERGQSLAQMALAVLREQEGYAGDTVTSALIGA			
	SSVQLDNLSDLSLNNLEFSDAELEAIDISHDGINIWAATDSKTREN"			
BASE COUNT	526 a	640 c	648 g	560 t
ORIGIN				
Query Match	100.0%;	Score 711;	DB 2;	Length 2374;
Best Local Similarity	100.0%;	Pred. No. 2.7e-207;		
Matches 711;	Conservative	0;	Mismatches	0;
			Indels	0;
			Gaps	0;
Qy	1	gtgtcatcgccgacggttgcttcttcaatgatctgtggcgcttccacctgtttt	60	
Db	1359	GTGTATCGGGCAGCGGTGCTGTTCTTCTTCAATGATCTGTGGCGCTTCCACCTGTTT	1300	
Qy	61	gtcatggcgctcttcgctgccgacgacgcaaacaccataacagtgatgcaccccgag	120	
Db	1299	GTGATGGCGTCTTCGTCGCATGACGGCAACACCAACAGGTAAGCGGATGCCACCCAG	1240	
Qy	121	cgcataataatcagacacgacgacgacgacgacgacgacgacgacgacgacgacg	180	
Db	1239	CGCATAAATATCAGCAGCATCGCGCGGCGGATTTGGACAAAAGATCAACGCCCGG	1180	
Qy	181	gcgatgaacaaagacgctcagaataataaacacacgacgacgacgacgacgacgacg	240	
Db	1179	GGGATGAACAAAAGACGTCAGAAATTAACACACGAGAACCCCATGATGCTCTCG	1120	
Qy	241	cgttaattccttggtttaataaccagtagtacattctcggtcccgatggacagtagtaag	300	
Db	1119	CGCTTAATTCCTTGTGTTAATACACAGTACATCTCTCGGTCGATGGACAGTAAAG	1060	
Qy	301	gcccccaagcagacacctgtaataagaatttcacattccacattcgtagcattggaat	360	

```
Db 1059 GCCCCAAAAGCAGACCTGTAATGAGATTTCCATGATCACCATCGTGACCTATGGAAGT 1000
Qy 361 acttaagtaaaatgattggttttaacatggttttaataatagcttcaatgaacccattcaaa 420
Db 999 ACTTAAGTAAATGATGTTGTTTAAACATGTTTAAATATAGCTTCATGAACCCCAATTCAA 940
Qy 421 ctgacacattgctctcaatcatgatgaagcagcttcgaagcgccctcttagccctt 480
Db 939 CTGGACATTTGCTCTCAATCATGATGAAGCGAGCTTCGAAGGCGCTCCTTAGCCCTT 880
Qy 481 tccatttccccctcggggtgagtcagcgcttaaaagctctcgcagcatcacagtggtgca 540
Db 879 TCCATTTCCCTCGCGGTGAGTCAGCGCTTAAAGCTCTCGACATCATCGTGGTGA 820
Qy 541 gtgttgatcgccacccacccacccacccacccacccacccacccacccacccacccaccc 600
Db 819 GTGTTGGTATCGCGCACCACCGCCAAAGCAACCGAAGCGGTGAAGTCTTGTGCAA 760
Qy 601 gcagcgagaaatggttctgcaagcagaaactaaagcgaactatctggagccctt 660
Db 759 GCAGCGGGAATGTTGCTGCAAGCAGAACTAAAGCGCAACTATCTCGACGCTT 700
Qy 661 gctgaatcccgcttaaccatcgccatcaacgagatctgctatccacatgg 711
Db 699 CCGAATCCCGTTAACCATCCCATCCCATCAACGACATTCGCTATCCACATGG 649

RESULT 2
A93933/c
LOCUS A93933 2374 bp DNA PAT 22-JAN-2000
DEFINITION Sequence 2 from Patent WO9723597.
ACCESSION A93933
VERSION A93933.1 GI:6742037
KEYWORDS
SOURCE
ORGANISM Corynebacterium glutamicum.
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Corynebacterineae; Corynebacteriaceae;
Corynebacterium.
REFERENCE
1 (bases 1 to 2374)
AUTHORS Vrljic,M. and Eggeing,L.
TITLE PROCESS FOR THE MICROBIAL PRODUCTION OF AMINO ACIDS BY BOOSTED
ACTIVITY OF EXPORT CARRIERS
JOURNAL Patent: WO 9723597-A 2 03-JUL-1997;
KERNFORSCHUNGSANLAGE JUELICH (DE); VRLJIC MARINA (DE)
FEATURES
Location/Qualifiers
source 1. .2374
/organism="Corynebacterium glutamicum"
/db_xref="taxon:1718"
BASE COUNT 526 a 640 c 648 g 560 t
ORIGIN

Query Match 100.0%; Score 711; DB 9; Length 2374;
Best Local Similarity 100.0%; Pred. No. 2.7e-207;
Matches 711; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gtgtcatcgggcaggtggtggttcttcaatgattggtggtggttccacattgttt 60
Db 1359 GTGTCATCGGCGAGGTGTTGTTCTTCATGATCTGTGCGCTTCCACCTGTTT 1300
Qy 61 gtaatggtcttctgctgcatgacggcgaacacacacacacacacacacacacacacacac 120
Db 1299 GTCATGCGCTTTCGCTGTCATGACGCGCAACCATCAACAGGTAAGCGATGCCACCCAG 1240
Qy 121 cgcataatcgcagcagatcgccggtgattgacaaaagatacaacgccaaggtaccg 180
Db 1239 CGCATATATCGACCATCGCGCGCGCAATGGACAAAAGATCAACGCCAAGGTGCGG 1180
Qy 181 gcgatgaacaaaagcgtcagaaattaaacacacacacacacacacacacacacacacac 240
Db 1179 GCGATGAACAAAAGCGTCAGAAATTAACACACACACACACACACACACACACACACAC 1120
```

```
Qy 241 cgttaattccttgtttaataccacagtacattctcgggtccgatggacagtaaaagactg 300
Db 1119 CGCTAATTCCTGTTTAAATCACCAGTACATFTCTCGGTCCGATGGAGTAAAGAACTG 1060
Qy 301 gccccaaaagcagacacctgtaataagatttccatgatcacccatcgtagcctatggaagt 360
Db 1059 GCCCCAAAAGCAGACCTGTAATGAAGATTTCCATGATCACCATCGTGACCTATGGAAGT 1000
Qy 361 acttaagtaaaatgattggttcttaacatggttttaataatagcttcaatgaacccattcaaa 420
Db 999 ACTTAAGTAAATGATGTTGTTTAAACATGTTTAAATATAGCTTCATGAACCCCAATTCAA 940
Qy 421 ctgacacattgctctcaatcatgatgaagcagcttcgaagcgccctcttagccctt 480
Db 939 CTGGACATTTGCTCTCAATCATGATGAAGCGAGCTTCGAAGGCGCTCCTTAGCCCTT 880
Qy 481 tccatttccccctcggggtgagtcagcgcttaaaagctctcgcagcatcacagtggtgca 540
Db 879 TCCATTTCCCTCGCGGTGAGTCAGCGCTTAAAGCTCTCGACATCATCGTGGTGA 820
Qy 541 gtgttgatcgccacccacccacccacccacccacccacccacccacccacccacccaccc 600
Db 819 GTGTTGGTATCGCGCACCACCGCCAAAGCAACCGAAGCGGTGAAGTCTTGTGCAA 760
Qy 601 gcagcgagaaatggttctgcaagcagaaactaaagcgaactatctggagccctt 660
Db 759 GCAGCGGGAATGTTGCTGCAAGCAGAACTAAAGCGCAACTATCTCGACGCTT 700
Qy 661 gctgaatcccgcttaaccatcgccatcaacgagatctgctatccacatgg 711
Db 699 CCGAATCCCGTTAACCATCCCATCCCATCAACGACATTCGCTATCCACATGG 649

RESULT 3
AX063771/c
LOCUS AX063771 822 bp DNA PAT 24-JAN-2001
DEFINITION Sequence 53 from Patent WO0100843.
ACCESSION AX063771
VERSION AX063771.1 GI:12541483
KEYWORDS
SOURCE
ORGANISM Corynebacterium glutamicum.
Corynebacterium glutamicum
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Corynebacterineae; Corynebacteriaceae;
Corynebacterium.
REFERENCE
1 (bases 1 to 822)
AUTHORS Pompeius,M., Kroeger,B., Schroeder,H., Zelder,O. and Haberhauer,G.
TITLE corynebacterium glutamicum genes encoding metabolic pathway
JOURNAL Patent: WO 0100843-A 53 04-JAN-2001;
BASF AKTIENGESELLSCHAFT (DE)
FEATURES
Location/Qualifiers
source 1. .822
/organism="Corynebacterium glutamicum"
/db_xref="taxon:1718"
CDS
101..802
/notes="RXA01394"
/codon_start=1
/transl_table=11
/protein_id="CAC25127.1"
/db_xref="GI:12541484"
/translation="MEIFITGLLLGLASLLLSIGPQNLVKGKIKREGLIIVLLVCLI
SDVFLFAGTLGVLLSNAPIVLDIMRWGGTAYLLWFAVMAAKADTKVAPQIE
EFTVPDDTPLGGSVAVDTRNRVREVSVDKRVWVPMMLMAIVLTWLPNAYLDA
FVFIGVGAQYDGTGRWIFFAAGAFASLIWFLVGFCAALSRLPSLPKVRWVWV
AVVMTALATKLMLMG"
BASE COUNT 167 a 192 c 246 g 217 t
ORIGIN

Query Match 61.2%; Score 435; DB 9; Length 822;
Best Local Similarity 100.0%; Pred. No. 1.6e-122;
Matches 435; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

QY	1	gtgtcatcgggcaaggttggtctgttcttcgaatgatctgttgcgcttcacactgttt	60
Db	435	GTGTCATCGGGCAGCGTTGGTCTGTCTTCAATGATCTGTGGCGCTTCACACTTGTT	376
QY	61	gtcatggcgctcttcgtctgcattgacggcaaacataacaggttaagcgatgccaccocag	120
Db	375	GTCATGGCGCTTTCGCTGCCATGACGGCAAAACCATTAACAGGTAAGCGATGCCACCOCAG	316
QY	121	cgcataatatcgagcacgatacggcgcgccattggacaaaagatacaacgccaaagtgccg	180
Db	315	CGCATAATATCGAGCACGATCGCGCGGCATTGGACAAAAAGATCAACGCCAAGGTGCCG	256
QY	181	gcgatgaacaaaaagcgtcagaataataacacacgagaagaacccaatlgagtcccttcg	240
Db	255	CGGATGAACAAAAGAAGCGTCAGAAATTAACACACGAGAGAACCGCCAATGAGTCCCTCG	196
QY	241	cgcttaattccttgtttaatacccagtcacattctgcggtccgcatggacagtaaagaactg	300
Db	195	CGCTTAATCTCTGTTTAAATCACCAGTACATCTCGCGGTCGGATGGACAGTAAAGAAGCTG	136
QY	301	gccccaaaaagcagacctgaatgaagatttccatgatcaccatcgtgacctatgggaagt	360
Db	135	GCCCCAAAAGCAGACCTGTAATGAAGATTTCATGATCACCATCGTAGCTATGGAAGT	76
QY	361	acttaagtaaaatgattgggtcttaacatggtttaatatagcttcatgaaccccatcaa	420
Db	75	ACTTAAGTAAATGATTGGTCTTTAACATGGTTTAATATAGCTTCATGAACCCCATTCAA	16
QY	421	ctggacactttgctc	435
Db	15	CTGACACACTTTGCTC	1

```

RESULT      4
AX063767
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source
CDS
BASE COUNT
ORIGIN

```

Query Match 57.1%; Score 406; DB 9; Length 993;

[illegible]

RESULT	5
SCSF8/c	
LOCUS	SC5F8 bp DNA BCT 09-JUN-2000
DEFINITION	Streptomyces coelicolor cosmid 5F8.
ACCESSION	AL357613
VERSION	AL357613.1 GI:8347023
KEYWORDS	amino acid decarboxylase; anti-sigma factor; integral membrane protein; LuxR-family regulator; LysR-family transcriptional regulator; membrane transport protein; oxidoreductase; regulatory protein; RNA polymerase sigma factor; serine/threonine protein kinase; stress-inducible protein; TetR-family transcriptional regulator; thioredoxin reductase; threonine dehydratase; two-component sensor histidine kinase.
SOURCE	Streptomyces coelicolor A3(2).
ORGANISM	Streptomyces coelicolor A3(2) Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces. 1 (bases 1 to 33285)
REFERENCE	Redenbach,M., Kieser,H.M., Denapait,D., Eichner,A., Cullum,J., Kinashi,H. and Hopwood,D.A.
AUTHORS	A set of ordered cosmids and a detailed genetic and physical map for the 8 Mb Streptomyces coelicolor A3(2) chromosome Mol. Microbiol. 21 (1), 77-96 (1996)
TITLE	97000351
JOURNAL	2 (bases 1 to 33285)
MEDLINE	Seeger,K.J. and Harris,D.
REFERENCE	Unpublished
AUTHORS	3 (bases 1 to 33285)
REFERENCE	Thomson,N.R., Parkhill,J., Barrell,B.G. and Rajandream,M.A.
AUTHORS	Direct Submission
TITLE	Submitted (05-JUN-2000) Streptomyces coelicolor sequencing project,
JOURNAL	Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA E-mail: barrell@sanger.ac.uk Cosmids supplied by Prof. David A. Hopwood, [3] John Innes Centre, Norwich Research Park, Colney, Norwich, Norfolk NR4 7UH, UK
COMMENT	Notes: Streptomyces coelicolor sequencing at The Sanger Centre is funded by the BBSRC and Beowulf Genomics details of S. coelicolor sequencing at the Sanger Centre are available on the World Wide Web. URL: http://www.sanger.ac.uk/Projects/S_coelicolor/

COMMENT

CDS are numbered using the following system eg SC7B7.01c. SC (S. coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary strand).

The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous.

The length in codons is given for each CDS. Usually the highest scoring match found by fasta -o is given for CDS which show significant similarity to other CDS in the database. The position of possible ribosome binding site sequences are given where these have been used to deduce the initiation codon.

Gene prediction is based on positional base preference in codons using a specially developed Hidden Markov Model (Krogh et al., Nucleic Acids Research, 22(22):4768-4778(1994)) and the FramePlot program of Bibb et al., Gene 30:157-66(1994) as implemented at <http://www.nih.go.jp/jun/cgi-bin/frameplot.pl>.

CAUTION: We may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtg, ttg or att) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.

IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions.

Cosmid 5F8 lies on genomic restriction fragment AseI-A bordered by cosmids 5H1 and 4G10.

FEATURES

```

source
1. .33285
/organism="Streptomyces coelicolor A3(2)"
/strain="A3(2)"
/db_xref="taxon:100226"
/clone="cosmid 5F8"
complement(1..344)
/genes="SC5F8.01c"
complement(1..108)
/genes="SC5F8.01c"
/notes="nominal overlap with cosmid SC5H1 between bases 1..108."
complement(1..344)
/genes="SC5F8.01c"
/notes="SC5F8.01c, possible serine/threonine protein kinase, partials CDS, len: 114 aa. Similar to many proposed Streptomyces coelicolor serine/threonine protein kinases e.g. TR:CAB82014(EMBL:AL161755) SC63.07 (717 aa), fasta scores opt: 172 z-score: 222.0 E(): 6.8e-05 33.0% identity in 103 aa overlap. Overlaps and extends into CDS SC5H1.01 on the adjoining cosmid."
/codon_start=1
/transl_table=11
/product="putative serine/threonine protein kinase (fragment)."
/protein_id="CAB93729.1"
/db_xref="GI:8347024"
/translation="MGRAHVSHELVAGRYLFEVYQRETRNVCHSGEDATTCRCLVLTRELPFGAGGAARRAPGRVITGETMASLCPGRIAPVLDVADVADGLMTVTEWVAGVPLGDLDRRG"
complement(492..1451)
/genes="SC5F8.02c"
complement(492..1451)
/genes="SC5F8.02c"
/notes="SC5F8.02c, possible threonine dehydratase, len: 319 aa. Similar to many including: Escherichia coli SW:THD2_ECOLI(M21312) threonine dehydratase catabolic (EC 4.2.1.16) TdCB (329 aa), fasta scores opt: 609 z-score: 676.6 E(): 3.2e-30 39.4% identity in 312 aa overlap. Contains a Prosite hit to Serine/threonine dehydratases pyridoxal-phosphate attachment site and a pfam match to entry PF00291 S-T.dehydratase, Pyridoxal-phosphate dependent enzyme."
/codon_start=1
/transl_table=11
/product="putative threonine dehydratase."

```

```

/protein_id="CAB93730.1"
/db_xref="GI:8347025"
/translation="MIAITEIAAAERIAHVVRTPVPSPGLSALLGVPVTKLELLQRTSGPKARGATAKLLSLTEAERAGAVVSGNGHIAVAVMAADLVKATVVMPTA PARSVIAEAGALVRLTDMDSAPALVTRLREEGTLVHPDFDPVVAAGQCTVGLGF AEDSLTDVLYISGGGLIAGVAALRALRPGVRVGVETEGAEAMSALAGSPLT PLSSVVTLSAFPSVSRILYDHVAELVTELVVPDREAVRGSALALAHAKVWTEPAAG CLLPAAARRVVERGVGRIGLVVCNGGNATVGDMAVWADRFLGR"
complement(546..1412)
/genes="SC5F8.02c"
/notes="Pfam match to entry PF00291 S-T.dehydratase, pyridoxal-phosphate dependent enzyme, score 210.10, E-value 3.3e-59"
complement(1287..1328)
/genes="SC5F8.02c"
/notes="PS00165 Serine/threonine dehydratases pyridoxal-phosphate attachment site"
complement(1477..2154)
/genes="SC5F8.03c"
complement(1477..2154)
/notes="SC5F8.03c, unknown, len: 225 aa. Similar to several proteins of undefined function including: Deinococcus radiodurans TR:Q9RWP1(EMBL:A5001920) conserved hypothetical protein (254 aa), fasta scores opt: 467 z-score: 555.6 E(): 1.8e-23 36.8% identity in 205 aa overlap and Streptomyces coelicolor TR:CAB70638(EMBL:AL137242) SC8F4.12C (239 aa), fasta scores opt: 397 z-score: 474.2 E(): 6.1e-19 36.1% identity in 183 aa overlap."
/codon_start=1
/transl_table=11
/product="hypothetical protein SC5F8.03c."
/protein_id="CAB93731.1"
/db_xref="GI:8347026"
/translation="MKGDLFSSEHMVQPARQMTVENSICIRYAVNGEMLARQGAMIATRNLQFERKQGVGMUKRAVTGEGPLMAVRQGGENAFANEQNFVVEFGDEFTVNGRNVLDCASUSYRIATVKGSIAGGLFNSVFTGGQRLGLVCENPLVIPVSQYQPVHVDITDVAVGVSGAGTSLRSQSIGSMLRGGSGEAVQLVLQGGFVVVVRPEATPOKPOQH"
complement(2166..3470)
/genes="SC5F8.04c"
complement(2166..3470)
/genes="SC5F8.04c"
/notes="SC5F8.04c, unknown, len: 434 aa. Similar to Streptomyces coelicolor TR:Q9XW9(EMBL:AL096743) putative transport associated protein, SC17.21 (399 aa), fasta scores opt: 428 z-score: 449.2 E(): 1.5e-17 38.3% identity in 209 aa overlap."
/codon_start=1
/transl_table=11
/product="hypothetical protein SC5F8.04c."
/protein_id="CAB93732.1"
/db_xref="GI:8347027"
/translation="MSDLVPGGNVPLPGGVSVRVPGGFDVSLVTDGGKVGSGDAFVFYNOPEAGARLDDTLVDPARLRGAARVTVAVGSDGTPGLGALPSPVLVTDARGTVARTPARPGRTEVLLVFYRGGEGWKVRALGOGYADGLAGLDFGVEYDIDAPPEPATPEPTPGADSDGLNLSNRAAAGSPAVRDPRLRAAHAAHMAAAGTLSIETRDGVSVHVRVSGAVLTVGHELVSGPRTPAEFVAYCLRAERTRLTLDHTAETHAGWACVTGPGSDTYVTALWAVPLTPDLARTTAEVVGLTNRERAGAGLPALAVDARLTAQAQHSADMYTRDFYSHTDPDGCKPWRRAAAGADRSVGENTACGQSPAEVVEGMNSPGHRAINLEAGFTHIGVLAGGGRAGTWTQLLG"
complement(3562..6528)
/genes="SC5F8.05c"
complement(3562..6528)
/genes="SC5F8.05c"
/notes="SC5F8.05c, possible LuxR-family regulator, len: 988 aa. Similar to several other putative regulators from Streptomyces coelicolor including: TR:Q9XAN4(EMBL:AL079355) SC4C6.06 (893 aa), fasta scores opt: 305 z-score: 328.7 E(): 7.7e-11 26.4% identity in 910 aa overlap and TR:Q92573(EMBL:AL035569) SC8P9.18 (1091 aa), fasta scores opt: 857 z-score: 926.1 E(): 0.36.5%

```

identity in 1077 aa overlap. Contains a Prosite hit to PS0017 ATP/GTP-binding site motif A (P-loop) Pfam match to entry PF00196 GerE, Bacterial regulatory proteins, luxR family."

```
/codon_start=1
/transl_table=11
/product="putative LuxR-family regulator."
/protein_id="CAB93733.1"
/db_xref="GI:8347028"
/translation="MGVPTQSLTSPVGRDELARLGVLEARAGARAVLVAGDA
GVKTRTLEHAGRAAGTTLTGCHVDLDGVLGYLPFTTEILCVLAADERFAAYLA
GVNADRLGGPDGDTDAEPLRLEFVLAALTELDAPLLVLEDLHWQSSR
DLRLGLRGVLPAGGAPGRALFASVADDLHRRHRLPLLAELVLPQVVERLE
LRPDSVAVLRRLRPLRPSVTRIVERAGNFAEELVAATDAPAGVPSGL
ADVLRIEQLSETAQVLRTAAGVRVGHDLRLDVAVLPEEELSAVERQLL
```

Query Match 11.4%; Score 81.2; DB 3; Length 33285;

Best Local Similarity 59.6%; Pred. No. 2.1e-13;

Matches 137; Conservative 0; Mismatches 93; Indels 0; Gaps 0;

QY 418 caactggacattgtctcaatcattgatgaagcagcttcgaagcgccctccttagcc 477

DB 18092 CAGGTGCGGACCTGCTGCGCGTGTGGAGAGGACACCTTCACGCGCGCGCGCGCC 18033

QY 478 ctttcattccccctggcggtgagtcagcgcttaaaagctctgagcatcacgtgggt 537

DB 18032 CTGACGTCAGCGCGCGGTGTCAGCCAGCGGGTGAAGGCACGTGAGCAGCGCACCGGC 17973

QY 538 cgaagtgttgatcgacacccacacgacgaacacgacgagcggtggaacgttggt 597

DB 17972 CGGGTGTCTGCTGCGGACCAAGCGGTGCGGCGGACGATTCGCGCGCGCGTGTG 17913

QY 598 caagcagcgcaaaatgtgtgtctgcaagcagaaactaaagcgcaact 647

DB 17912 CGGTGCGCGCGGAGTGTGCGGCGTGTGAGCGGACGCGCTCGCGCGAGCT 17863

RESULT 6

AB011413/c 12070 bp DNA BCT 07-AUG-1998
LOCUS
DEFINITION
Streptomyces griseus genes for Orf2, Orf3, Orf4, Orf5, AfsA, Orf8,
partial and complete cds.

AB011413
ACCESSION
VERSION
KEYWORDS
SOURCE

AB011413.1 GI:3401946
Orf8; AfsA; Orf5; Orf4; Orf3; Orf2.

Streptomyces griseus DNA.

Streptomyces griseus

Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;

Actinomycetales; Streptomyceae; Streptomycetaceae; Streptomyces.

1 (bases 1 to 12070)

Umeyama, T.

Direct Submission

Submitted (23-FEB-1998) to the DDBJ/EMBL/GenBank databases. Takashi

Umeyama, University of Tokyo, Department of Agriculture and Life

Sciences; Yayoi 1-1-1, Bunkyo-ku, Tokyo 113, Japan

(E-mail: aa67103@hongo.ecc.u-tokyo.ac.jp, Tel: +81-3-3812-2111)

2 (sites)

Umeyama, T.

Open reading frame encoded around afsA gene

Unpublished (1998)

Location/Qualifiers

1..12070

/organism="Streptomyces griseus"

/db_xref="taxon:1911"

complement(2660..3511)

/gene="orf2"

complement(2660..3511)

/gene="orf2"

/function="sensor histidine kinase"

/codon_start=1

/transl_table=11

/product="orf2"

/protein_id="BAA32130.1"

```
/db_xref="GI:3401947"
/translation="VAAPVPAAPHPGPEAAAPAAARLTAAYTLITVVGLACLISLVIR
TDDRARMDEMRRRASVASLIVYEDDIRGAPAEHRSLVLSIVIGAGALTALA
ALTGHLGSRGCGPAWEALFOERLADAAHELRTPVAVMRGSVEVAAAGAPGLEGL
PRIIRAADRMADVVENLITRGRLEAGRTSYGRSLRLDLQVVEECAELPEGGPGLER
LEESVVEADAALVRVAVNLDNNAVRHGGPSRGAAGRTAGDRVTRTGVRVADRGPAV
APARLPS"
```

complement(3523..>4197)

/function="two-component regulator protein"

/codon_start=1

/transl_table=11

/product="Orf3"

/protein_id="BAA32131.1"

/db_xref="GI:3401948"

```
/translation="MRVLLVEDDDLDVIGAGLRDGGFAMDCASDMPDADVLLHLSA
YCVVLDPMVPSGDTLAPLEGRNRAGNSVPVLCITALDLSLDERLGLSGADVDYLAKP
FSMRELVLVRGLRRASRLSFLGCADVYMDVARHEVRREGVLLSLSPREYAVLQV
LLVHRDVTVTGTGLEHCWDEMDPVSVDVAVAGLRKRLGSPGLVHTVRGQGFLLS
ADPGPS"
```

complement(4333..>5250)

/function="methyltransferase"

/codon_start=1

/transl_table=11

/product="Orf4"

/protein_id="BAA32132.1"

/db_xref="GI:3401949"

```
/translation="MATCALCTALRGVDFDRIGTGLRAELAGSLGTHPRATLRLLR
ALAGLQLSEPEPCAFRTTAAAGNVLRADAPCTVMAMRFTDPMVLRGMDLLDSVFT
GETTDTVGTDFDFFHREHPELSAFNEASQGTRLTAETVPHHYDGRQRRLVDIG
GDDTTLASILRAHOEPGRVLFDTAAGLAQAPRLAREGLDGRVLTGDFPFASAPAG
GDXLINSIIHWDVDAVRHLRHIDVTPDHGSLLLIVEPLVLPATVPADRPDNNVLSDL
NMLVNVGGRTADDFAALCTAGGFACGA"
```

complement(5450..6361)

/function="regulator protein"

/codon_start=1

/transl_table=11

/product="Orf5"

/protein_id="BAA32133.1"

/db_xref="GI:3401950"

```
/translation="MSELPDQVRLAVVDEGTFDSATLRLTPSAVSORVKALEQ
RTGVLLNRTKVPRTSEGEVVALARLARLEHEAQAALGMSGPEPTLLPIVNSDS
LATWSQPCGACRRSDSATPAGPGPYGRPAAGGADGDLVAGGCGDLFGPAG
RMRLPVASPGFADRWLGRDGTALRELIGAPVCFDRDDLDQADFVRRUGPGARPS
ARRHLVPTSEGFANAVSGMGWVPEYQAEPLLSGDLRLVLAPEPTVDVPLYQOQWK
LESPALAAVAEAAAEAAEALDEAPPN"
```

7486..8391

/codon_start=1

/transl_table=11

/product="AfsA"

/protein_id="BAA32134.1"

/db_xref="GI:3401951"

```
/translation="MDAEVVPVGVIEVHRTPEDAFPRNVRLGRDRFAEAVLP
HDHFFAPVGGDLHDPLLVAAEQMAALAFHAGYGIPLGYHLLTLDYVCHPEHLG
VGGPTEIGLVFCSDLKWRAGLPAQGRVNAVHKGDLAATGAATRFSTPKAYRRM
RGDVPVEIGSILPETAVPASPGARVEDVVLSTGREGVWELRVTRHPTLFRPND
HVPGLMLEAARQAACLVAGPAGVIVPEARTFRHRYSEFGSPCWIGAVVQCADEDTV
TVRVTHGODGETVFSTVLSGPRANG"
```

10558..11598

/function="alcohol dehydrogenase"

/codon_start=1

/transl_table=11

/product="Orf8"

/protein_id="BAA32135.1"

/db_xref="GI:3401952"

```
/translation="MTTVAAVAAPRAKAPLERTTIERPVGFDILIDIKFAGICHSD
THQARDWGEGIFPMVPGHEIAGVATGPGVTRFKYGDVRGVCMDSCCTCDACL
GREQHCAGNTQTNALDRNGEPTYGGSTHLVTEKFAISIPGIALDEAPPLICAG
ITTYSPRLRWGAGKGVAVVGLGGLGHMAVIAHAUGAEVTVLSQSKKDDGLKIG
ADHYATGDPRTFELAGTFDVILSTPADFGAYLALLTEGLTVNVPPEPVS
NLFSLILGNRSIAGSAGIGIETQEMLDFAFCAGLGAIEVIGAQVNEAYERVLS
VYRFVDTATI"
```

1647 a 4478 c 4272 g 1673 t

BASE COUNT
ORIGIN

AUTHORS	Mahairas,G., Sabo,P.J., Hickey,M.J., Singh,D.C. and Stover,C.K.
TITLE	Direct Submision
JOURNAL	Submitted (24-AUG-1995) Mark J. Hickey, Molecular Micro., Pathogenesis Corp., 201 Elliott Ave. W., Seattle, WA 98119, USA
FEATURES	Location/Qualifiers
SOURCE	1. 14844
repeat_unit	/organism="Mycobacterium bovis" /db_xref="taxon:1765" 2864..2888
misc_difference	/note="One of two imperfect direct repeats located at or near the deletion breakpoints"
gene	/note="this interval is completely absent in the closely related, but avirulent strain Mycobacterium bovis BCG"
CDS	10705..11391 /gene="mpt64" 10705..11391 /gene="mpt64" /note="immunogenic protein" /codon_start=1 /transl_table=11 /product="MPT64" /protein_id="AAC44034.1" /db_xref="GI:1049225"
repeat_unit	13652..13675 /note="one of two imperfect direct repeats located at or near the deletion breakpoints"
BASE COUNT	2800 a 4682 c 4603 g 2759 t
ORIGIN	
Query Match	8.9%; Score 63.2; DB 3; Length 14844;
Best Local Similarity	55.3%; Pred. No. 6e-08;
Matches 166; Conservative	0; Mismatches 128; Indels 6; Gaps 2;
QY 418	caactgacacttgcctcaatcattgaagcagcttcgaagcgccctccttagcc 477
Db 4859	CAGCTGCCCGCATTTGCTGCCGTGGTTCGAACCTGGGCGAGCTTCGATGGCGCGGAGCGC 4918
QY 478	ctttccatttccccctcgccggtgagtcagcgcttaagctctcgagcatcacgtgggt 537
Db 4919	CTCATGTACACCCCTCGGCTCTCAGTCAGCGCATCAAGTCGTTGGAGCAGCAGTCGCG 4978
QY 538	cgagtggttgatcgccgaccacccgcccgaagcaacccgaagcggtggaagtccttg 597
Db 4979	CAGGTGCTGGTGTGTCAGGGAAGGCATCTCGGCGAGCGAGGTATCCCGCTGTTG 5038
QY 598	caagcagcggaataatggttgctgcaagcaaaataaagcaactatctg--ac 655
Db 5039	CGGTTGGCGCGCAACACGCTTGTCTCGAGTCCGAGCGGCTCGCTGAAATGGGTGGCAAC 5098
QY 656	gccttgctgaa----atcccggttaaccatcgccatcaacgagatgctgctatccatc 711
Db 5099	CGCTCGCTGAACGCGACGCGGATCACCATTGGGGTAACGCGGATTCATCGGCACATGG 5158
RESULT 8	
LOCUS	186263 15239 bp DNA PAT 10-JUN-1998
DEFINITION	Sequence 17 from patent US 5700683.
ACCESSION	186263
VERSION	186263.1 GI:3205981
KEYWORDS	
SOURCE	Unknown.
ORGANISM	Unknown.
REFERENCE	1 (bases 1 to 15239)
AUTHORS	Stover,C.Kendall and Mahairas,G.G.
TITLE	
JOURNAL	
MEDLINE	
REFERENCE	2 (bases 1 to 14844)
AUTHORS	Mahairas,G.G., Sabo,P.J., Hickey,M.J., Singh,D.C. and Stover,C.K.
TITLE	Molecular analysis of genetic differences between Mycobacterium bovis BCG and virulent M. bovis
JOURNAL	J. Bacteriol. 178 (5), 1274-1282 (1996)
MEDLINE	96200095
REFERENCE	3 (bases 1 to 14844)
AUTHORS	Oettinger,T. and Andersen,A.B.
TITLE	Cloning and B-cell-epitope mapping of MPT64 from Mycobacterium tuberculosis H37Rv
JOURNAL	Infect. Immun. 62 (5), 2058-2064 (1994)
MEDLINE	94222581
REFERENCE	4 (bases 1 to 14844)
AUTHORS	Mahairas,G.G., Sabo,P.J., Hickey,M.J., Singh,D.C. and Stover,C.K.
TITLE	Molecular analysis of genetic differences between Mycobacterium bovis BCG and virulent M. bovis
JOURNAL	J. Bacteriol. 178 (5), 1274-1282 (1996)
MEDLINE	96200095
REFERENCE	3 (bases 1 to 14844)

```

TITLE      Virulence-attenuating genetic deletions deleted from mycobacterium
JOURNAL    BCG
PATENT     Patent: US 5700683-A 17 23-DEC-1997;
FEATURES   Location/Qualifiers
Source     1..15239
           /organism="unknown"
BASE COUNT 2878 a 4791 c 4729 g 2841 t
ORIGIN

Query Match      8.9%; Score 63.2; DB 10; Length 15239;
Best Local Similarity 55.3%; Pred. No. 6.le-08;
Matches 166; Conservative 0; Mismatches 128; Indels 6; Gaps 2;

QY 418 caactggacacttgctctcaatcattgatgaaggcagcttcgaagcgccctcttagcc 477
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5253 CAGCTGGCGCATGGCTGCGTGGTCGAACTGGGCGAGCTTCGATCGCGCGGGAGCGC 5312

QY 478 ctttcatttccccctcgcggtgagtcgagcggttaaaagctctcagacatcacgtgggt 537
   || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5313 CTACATGTCACCCCTCGCTGTCAGTCAGCGCATCAAGTCGTTGAGCAGCAGGTCGCG 5372

QY 538 cgaagtgttgatcgcgcccccaacgcgcgaagcaaacgagcggtggaatccttg 597
   || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5373 CAGGTCTGCTGGTCTAGGGAAGAGCATGCTCGGGCGACGACCGCATCCGCTGTTG 5432

QY 598 caagcagcgcggaatatgtgtgtcgcagcagcaaaactaaagcgcaactatctg--ac 655
   || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5433 CGTTCGCGCGCGCAACACAGCGTGTCTGAGTCGAGTCGAGCGCGCTGCTGAATGGTGGCAAC 5492

QY 656 gccttgtaa-----atcccgtaaacattgcacatcaacgacgactcgcttccacatgg 711
   || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5493 GCCTGCTGAACGCGACCGGGATCACCATTCGCGTAAACGCCCATTCCTCATGGCAGATGG 5552

RESULT      9
AE005522    AE005522 10701 bp DNA BCT 21-MAR-2001
LOCUS       Escherichia coli O157:H7 EDL933 genome, contig 3 of 3, section 141
of 290.
ACCESSION   AE005522 AE005174
VERSION     AE005522.1 GI:12517448
KEYWORDS    Escherichia coli O157:H7 EDL933.
SOURCE      Bacteria; proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1 (bases 1 to 10701)
AUTHORS    Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
            Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
            Postai,G., Hackett,J., Link,S., Boutin,A., Shao,Y., Miller,L.,
            Grobeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamouis,K.,
            Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
            Welch,R.A. and Blattner,F.R.
            Genome sequence of enterohaemorrhagic Escherichia coli O157:H7
            Nature 409 (6819), 529-533 (2001)
            21074935
            11206551
REFERENCE   2 (bases 1 to 10701)
AUTHORS    Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
            Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
            Postai,G., Hackett,J., Link,S., Boutin,A., Shao,Y., Miller,L.,
            Grobeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamouis,K.,
            Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
            Welch,R.A. and Blattner,F.R.
            Direct Submission
            Submitted (22-OCT-2000) Laboratory of Genetics, University of
            Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
            Location/Qualifiers
            1..10701
            /organism="Escherichia coli O157:H7 EDL933"
            /strain="EDL933"
            /serotype="O157:H7"

/db_xref="taxon:155864"
/note="enterohaemorrhagic"
133..2277
/gene="sbm"
/note="24254"
133..2277
/gene="sbm"
/function="enzyme; Degradation of small molecules: Fatty acids"
/note="Residues 1 to 714 of 714 are 99.15 pct identical to residues 1 to 714 of 714 from Escherichia coli K-12 Strain MG1655: B2917"
/codon_start=1
/transl_table=11
/product="methylmalonyl-CoA mutase (MCM)"
/protein_id="AAG58043.1"
/db_xref="GI:12517449"
/translacion="MSNEQEWQOLANKELSRREKTVDSLVQQTABGIAIKPLYTEADL
DNLEVTGLPGLPYPRVATMTAQPTIRQVAGFSTAKESNAFYRNLAAGOKGL
SVAFDLATHRGVSDNPRVAGDVGTVAIDTVDKVLFDQIPDKMSVMTWNGAV
LPVLAFTIAAEQGVDPKLTGTIONDILKEYLCRNTYIYPKPSMRIADITAMCS
GNMPFRWTSISGYHMGAGANCYQVAFIADGIEYIKAAISAGLKIDDPRLSFF
FGIGMDLFMNVAMLRARYLWSEAVSGGGAQPKSLADKTHCQTSGWSLTQDPPYNNV
IRTTTEALATLGCTQSHNTAFDEALGLTDFDSARIARTQIILQIESESLCTVDP
AGSYIESITDQIVQARAIQQIDEAGGMAKIEAGLPKRMIEEASARESLIDQK
RVIVGVNKKYKLDHEDETDVLTEIDNVVNEQIASLERIRATRDDAAVTAALNATHAA
QHNNLLAAVNAARVRATLGEISDALEAFDRYLVPSQCVTVGIAQSYHOSEKSAE
FDAVAOTEFLADNGRPRILIAKMGLDGHDGAKVTASAYSDLGFDVDSLSPWETP
BEIARLAVENDVHVVGASSLAAGHKTLIPELVEALKKWGREDCICWAGGVIPPODYAF
LOERGVAAIYGGTPMLDSVRDVLNLISQHH"
2270..3265
/gene="ygfd"
/note="24255"
2270..3265
/gene="ygfd"
/function="phenotype; Not classified"
/note="Residues 1 to 331 of 331 are 99.69 pct identical to residues 1 to 331 of 331 from Escherichia coli K-12 Strain MG1655: B2918"
/codon_start=1
/transl_table=11
/product="putative nucleotide-binding protein"
/protein_id="AAG58044.1"
/db_xref="GI:12517450"
/translacion="MINEATLAESIRRLRQGEHATLAQAMTLVESRHPHQAALSTQLL
DAIMPYCQNTLRGLVGTGPGAGKSTFLFAGMLLIREGLKVAIVADPSPVTVGSGIL
GDKTRMDLARAFAFIRPVSSGHILGCASORARELMLLCEAAGYDVVIVETVGVGOS
ETEVANWVDCFISIQIAGGDDDLQIKKGLMEVADLVINKDDGDNHNTVIAIRHMYE
SALHILRRKYDEWPVRVLTCSALEKRGIDEIWHALIDFKALTASGRQQVQQQSVSE
WLRKQTEEEVLNHLFANEDFDYRYQTLLAVNNNTLSPTGLRQLSEFIQOYFD"
3234..4061
/gene="ygfg"
/note="24256"
3234..4061
/gene="ygfg"
/function="putative enzyme; Not classified"
/note="Residues 1 to 275 of 275 are 100.00 pct identical to residues 1 to 275 of 275 from Escherichia coli K-12 Strain MG1655: B2919"
/codon_start=1
/transl_table=11
/product="putative enzyme"
/protein_id="AAG58045.1"
/db_xref="GI:12517451"
/translacion="MNLSSRRNLIKIFMSYQYVNVTVINKVAVIEFNYGRKLNALS
VFIDDLQALSDLNRPETRCILRAPSGSKVFSAGHDHLPVSGRDPVSLSDPDLROI
NTRDQKPKIISWEGSVGCAFEIMSSDLIIAASSTFSMTFNPVNLGVYPNLVGIH
NLTRDAGFHIVKELIFTASPIAQRALAVGLNHVVEVEELEDTLQMAHHSKAPL
AIAVTKELRVLGEAHTMNSDEFRIQGRRAVYDSEYQEGMAFLFKRKPNEVGH"
4085..5563
/gene="ygfh"
/note="24257"
4085..5563
```



```

/gene="ygfH"
/function="putative enzyme; Not classified"
/notes="Residues 1 to 492 of 492 are 99.39 pct identical to
residues 1 to 492 of 492 from Escherichia coli K-12 Strain
MG1655; B2920"
/codon_start=1
/transl_table=11
/product="putative coenzyme A transferase"
/protein_id="AAG58046.1"
/db_xref="GI:12517452"
/translation="METQWTRMTADAEAEIIOHNDMVAFSGFTPGSPKALPTAIAARR
ANQHEAKPKYQIURLITGASISAAADDDLSADAVSWRAPYOTSSGLRKKINGAVSF
VDLHSEVQAVNYGFGDIDAVIASALAPGRVWLTSGGIGNAPTLLRAKKVILE
LNHYHPRVLAEDIVIPGAPRRNSVISIFHANDRGVTRIVQIDPKIIVAVVETNLPD
AGNHLQNPQCOIADNVPTFLLEQMAHGRIPPEFLPLOSQGVGNINNAVMAKLGEP
EIPPFMYSEVQESVHLETKISGASASLTIISADSLRKTYDNMDYFASRIYLRP
QETSNNPEIIRLGVLTALNVGLEDYGHANSSTHIVAGVLMNGIGSGDFPFIYALSI
FMAPSIAKEGISTVVMCMGSHVDNHSVKVITTEQGIADLRGLSLPLQARTIIDNCA
HPMYRDYLHRYLENAPGGHTHDLSHVFDLHRNLIAATGSMWG"
complement(5878..6456)
/gene="24258"
complement(5878..6456)
/gene="24258"
/function="putative regulator; Not classified"
/notes="Residues 1 to 192 of 192 are 99.47 pct identical to
residues 6 to 197 of 303 from Escherichia coli K-12 Strain
MG1655; B2921"
/codon_start=1
/transl_table=11
/product="partial putative transcriptional regulator
LYSR-type"
/protein_id="AAG58047.1"
/db_xref="GI:12517453"
/translation="MDIFISKMRNFILLAOYNNIARAEEKIHMTASPFKSTAALEE
QYGTTLTRKNDLSIGKQVYQLFVYQLSAIDNEIHNSGRSREIVIGIDNT
QPTIIPDQLISLGDYKGVGTQAPVEFSENGVIDNLDRDLDFIISPHVSRVQELN
LTISELPLRLNPLVSRVREERQELLQELP"
complement(6623..7363)
/gene="yggE"
/notes="24259"
complement(6623..7363)
/gene="yggE"
/function="phenotype; Not classified"
/notes="Residues 1 to 246 of 246 are 100.00 pct identical
to residues 1 to 246 of 246 from Escherichia coli K-12
Strain MG1655; B2922"
/codon_start=1
/transl_table=11
/product="putative actin"
/protein_id="AAG58048.1"
/db_xref="GI:12517454"
/translation="MKFKVIALAALMGISGMAQAQNELPDGPHIVTSGTASVDVAPDI
ATLAEVNVAAKDAATAKKQADERVAQYISFLELNOIAKKDDISSANLRTQPDYVDG
KSLKGYRAVRVVEVTLRQDLKNSLLDGALKAGLNEIRSVSLGVAQPDYKDKARKA
AIDALHQEALQELANGFHRKLGPIVSVRYHSNTQSPMPVRMKADAPVSAQETTEQA
AIOFDQDVDFQLEPVDQOAPKTPAAQ"
complement(7456..8091)
/gene="yggA"
/notes="24260"
complement(7456..8091)
/gene="yggA"
/function="orf; Unknown function"
/notes="Residues 1 to 211 of 211 are 98.10 pct identical to
residues 1 to 211 of 211 from Escherichia coli K-12 Strain
MG1655; B2923"
/codon_start=1
/transl_table=11
/product="orf, hypothetical protein"
/protein_id="AAG58049.1"
/db_xref="GI:12517455"
/translation="MFSYVFOGLGALGAAMILPLGPONAFVMNOGIRROYHIMIALICA
ISDLVLICAGIFGGSALLMQSPWLLALVTWGGVFWLLWYGFCAFKPTMSSNIELASAE
VLKQGRWKIATMLAVTLNPHVYLDTFVVLGGQLDVBXPFRWFALGTISASFLWF

```

```

FGLAIIAANLAPRLRTAKSORIINLVVGCVMFIALQLARDGIAHAQALFS"
complement(8230..9090)
/gene="yggB"
/notes="24261"
complement(8230..9090)
/gene="yggB"
/function="putative transport; Not classified"
/notes="Residues 1 to 286 of 286 are 100.00 pct identical
to residues 1 to 286 of 286 from Escherichia coli K-12
Strain MG1655; B2924"
/codon_start=1
/transl_table=11
/product="putative transport protein"
/protein_id="AAG58050.1"
/db_xref="GI:12517456"
/translation="MEDLNVDVDSINGAGSWLVANQALLSYAVNIVAALAIIVGLII
ARMISNAVRLMISRLKIDVADFSLVRYGIIAFTLIAALGRGVQVTSVIAVLGA
AGLAVGLALQGLSLNLAAGVLLVMPFRPFRAGEVVDLGGVAGTVLSVQIFSTMTRTADG

Query Match      8.7%; Score 62; DB 1; Length 10701;
Best Local Similarity 51.4%; Pred. No. 1.3e-07;
Matches 143; Conservative 0; Mismatches 135; Indels 0; Gaps 0;

QY 49 tcacacctgttgcacgtctcttcgtcgccatgacggcaaacaccatacaggtacgcg 108
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 7802 TCAATATTACTATCATCGCTGTTTAAAGCGCCAAACCATACCACAGCAAGAGACT 7861

QY 109 atgccaccacgcgcataatcgagcagcgtcgcgcgccattggacaaaagatcaacy 168
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 7862 ACGCCGCCCGAGTGACCGCCAGCAGCACCGCGCGCTGCATCAATAACGCCGTACCA 7921

QY 169 cccaaaggtgcggcgatgaacaaaagacgctcagaataataaacacacagagaagacgcga 228
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 7922 CCAAAATCCCGCGCAATCAGGACCAATCGCTGATCGCGCAAAAGTAGGCAATCATA 7981

QY 229 atagtccttcgcgccttaattccttttaataaccagtcattcgcgtccgagtcgac 288
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 7982 ATGTGTTACTGACGAGATGCCCTGATTTCATCAAAAGCATTTTGTGGCGCGCGGC 8041

QY 289 agtaaaagactggtcccccaaaagcagacactgtaataagaa 326
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 8042 AGAATCATAGTGTCCCAAGCAAGACACTTGAAAGTA 8079

RESULT 10
AP002563 266658 bp DNA BCT 07-MAR-2001
LOCUS Escherichia coli O157:H7 DNA, complete genome, section 14/20.
DEFINITION AP002563 BA000007
ACCESSION AP002563.1 GI:13363121
VERSION
KEYWORDS
SOURCE Escherichia coli O157:H7 (strain:O157:H7, sub_strain:RMD 0509952)
DNA.
ORGANISM Escherichia coli O157:H7
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
REFERENCE 1 (sites)
AUTHORS Makino,K., Yokoyama,K., Kubota,Y., Yutsudo,C.H., Kimura,S.,
Kurokawa,K., Ishii,K., Hattori,M., Tatsuno,I., Abe,H., Iida,T.,
Yamamoto,K., Ohnishi,M., Hayashi,T., Yasunaga,T., Honda,T.,
Sasakawa,K., and Shinagawa,H.
Complete nucleotide sequence of the prophage VT2-Sakai carrying the
verotoxin 2 genes of the enterohemorrhagic Escherichia coli O157:H7
derived from the Sakai outbreak
Genes Genet. Syst. 74 (5), 227-239 (1999)
JOURNAL 20198780
MEDLINE
REFERENCE 2 (sites)
AUTHORS Ohnishi,M., Murata,T., Nakayama,K., Kuhara,S., Hattori,M.,
Kurokawa,K., Yasunaga,T., Yokoyama,K., Makino,K., Shinagawa,H. and
Hayashi,T.
Comparative analysis of the whole set of rRNA operons between an
enterohemorrhagic Escherichia coli O157:H7 Sakai strain and an
Escherichia coli K-12 strain MG1655

```


Squares, S., Sqaers, R., Sulston, J.E., Taylor, K., Whitehead, S. and Barrell, B.G.
Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence
Nature 393 (6685), 537-544 (1998)
98295987
Erratum: [[published erratum appears in Nature 1998 Nov 12:396(6707):190]]
12 (bases 1 to 38500)
Parkhill, J.
Direct Submission
Submitted (11-JUN-1998) Submitted on behalf of the Mycobacterium tuberculosis sequencing and mapping teams, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA Unite de Genetique Moleculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux, 75724 Paris Cedex 15, France E-mail: parkhill@sanger.ac.uk
On Jun 27, 1998 this sequence version replaced gi:1403436.

Notes:
Details of *M. tuberculosis* sequencing at the Sanger Centre are available on the World Wide Web.
<http://www.sanger.ac.uk/projects/M.tuberculosis/> CDS have been renumbered from the original cosmid submissions but the old gene designations are in brackets after the new gene numbers.
Gene prediction was based on a Hidden Markov Model of TB genes implemented in TBparse (Krogh) supplemented with visual inspection of positional base preference in codons, especially where there is an increase in the observed/expected third position G + C.
CAUTION: In some cases we may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gta, or ttg) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.

Location/Qualifiers
1..38500
/organism="Mycobacterium tuberculosis"
/strain="H37Rv"
/db_xref="taxon:1773"
<1..38419
/organism="Mycobacterium tuberculosis"
/strain="H37Rv"
/db_xref="taxon:1773"
/clone="Y39"
complement(91..777)
/gene="mpt64"
complement(91..777)
/gene="mpt64"
/note="RV1980C, {MTCY39.39}, len: 228. mpt64 identical to MP86_MYCB0 P19996 immunogenic protein mpb64/mpt64 (228 aa). Similar to RV3036c, {MTV012.51c}"
/codon_start=1

```

/product__mp04
/protein_id="CAA98382.1"
/db_xref="GI:1403475"
/db_xref="SWISS-PROT:P19996"
/translation="MRKIFMLVAVLCCSGGVATAKTPTEELKGTDTGQAOIQ
MSPDPAVINILSPSYVDQKSLNAV IACTRQKFLSAATSPTRPEALYNITSATYQS
AIPRGTOAVLVKLYQVGNAGTHPTTYYKAFDWDQAYRKPIYDTLWQADTDLPLPVVFP
IVQGEUSLKGQGGQVQSVIAPNAGLDPVNYFNFAVDGVIFFFPFGELLPSAAGTQTVLV
PSRAIDSMLA"
/complement(787..791)
/note="ggagg, potential rbs upstream of Rv1980c"
/complement(1968..1936)
/gene="nrdF"
/complement(1968..1936)
/gene="nrdF"
/note="Rv1981c, (MTCY39_38), ribonucleotide reductase
small subunit, len: 322, highly similar to R1R4_SALTY
P17424, ribonucleoside-diphosphate reductase (319 aa)
FASTA scores, opt: 1402, z-score: 1716.0, E(): 0, (66.08
identity in 315 aa overlap); Also similar to
Rv3048c, (MTV012.63c); containsPS00368 Ribonucleotide

```


REFERENCE	AUTHORS	TITLE	JOURNAL	FEATURES	source	CDS	ORIGIN
1 (bases 1 to 636)	Nakanishi, K., Aleshin, V.V., Livshits, V.A., Tokhmakova, I.L., Troshin, P.V. and Zakataeva, N.P.	Method for producing l-amino acids	Patent: EP 1016710-A 15 05-JUL-2000;	AJINOMOTO KK (JP)	Location/Qualifiers 1. .636 /organism="Escherichia coli" /db_xref="taxon:562"	1. .636 /note="unnamed protein product" /codon_start=1 /transl_table=11 /protein_id="CAC09317.1" /db_xref="GI:10190303" /translation="MFSYFQGLGALGAMILPLGPQNAFVWNGGIRROYHIMIALICA ISDLVLCAGIFGSAWLLQFSLALVTWGGVAFLLWYGFARFKTAMSNIELASAE VKGQRWKLITATMLAVTAWLPHVYDFTFVLGSGGOLDVEPKRWKFAIGTISASFLLWF FKALLAAWLPAIRLTAQAKRIINLVGVCVMFIALQALQARDGIAHAQALFS"	BASE COUNT 120 a 143 c 180 g 193 t
Query Match	Best Local Similarity	8.5%; Score 60.4; DB 9; Length 636;					
Matches 142; Conservative	0; Mismatches 136; Indels	0; Gaps	0;				
QY 49	tcacacctgtttgtcatgcgtcttttcgctgcatgacgagcaaacataacagtgtaagcg	108					
Db 290	TCAATATTACTGCTCATTCCTCTTTTAAAGCGCCAAACCATACACAGCAAGAAGCCT	231					
QY 109	atgccaccacgacgataatatcgagcagcatcgcgcgcttggtgacaaagatcaacg	168					
Db 230	ACCCGCCCGGCTGACACGCGCCGACCAACGCGGACTGCATCAATAACGCGCTGCCA	171					
QY 169	cccaagtgccgagcagatgacaaaagaagtcgagaaatataacacacagagaacccgca	228					
Db 170	CCAAAATCCCGCGCAATCAGGACCAATCGCTGATGATGACACAAAGTAAGGCAATCATA	111					
QY 229	atgagtccttcgcttaattcctgttttaaccacgatacatctcggtccgatggac	288					
Db 110	ATGTGTACTGACGACGTATGCTGCTGATTCATCAAAAGCATTTTGTGGACGCGGT	51					
QY 289	agtcaaaagactgcccccaaaagacagcagctgtaatgaa	326					
Db 50	AGCATCATAGCGCCCGCCAGTCAGACAGCTTGCAAGTA	13					
RESULT 13	ECFDPAGK/C						
LOCUS	8029 bp DNA BCT 21-MAR-1994						
DEFINITION	Escherichia coli fda, pgk and gapB genes for fructose 1,6-bisphosphate aldolase (class II), phosphoglycerate kinase and glyceraldehyde 3-phosphate dehydrogenase.						
ACCESSION	X14436.1						
VERSION	GI:41417						
KEYWORDS	fda gene; fructose 1,6 biphosphate aldolase; gapB gene; glyceraldehyde 3-phosphate dehydrogenase; pgk gene; phosphoglycerate kinase; unidentified reading frame.						
SOURCE	Escherichia coli.						
ORGANISM	Escherichia coli						
	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.						
REFERENCE	1 (bases 1 to 8029)						
AUTHORS	Alefounder, P.R.						
TITLE	Direct Submission						
JOURNAL	Submitted (22-FEB-1989) Alefounder P.R., University of Cambridge, Department of Chemistry, Lensfield Road, Cambridge CB2 1EW, England						
REFERENCE	2 (bases 4376 to 5508)						
AUTHORS	Alefounder, P.R., Baldwin, S.A., Perham, R.N. and Short, N.J.						
TITLE	Cloning, sequence analysis and over-expression of Escherichia coli class II fructose 1,6-bisphosphate aldolase of Escherichia coli						

```
CDS
3023..4186
/note="phosphoglycerate kinase (AA 1-387)"
/codon_start=1
/transl_table=11
/protein_id="CAA32604.1"
/db_xref="GI:41422"
/db_xref="SWISS-PROT:P11665"
/translacion="MSVIRKTDLDLAGKRVFIRADLNVPVKDKVTSIDARISLPTI
ELALOGKAVMTKGRVTEGEYNEEFSLLPVNYLKDLSNPVRLVKDYLDGVDVA
EGELVVLNVNRKKEPKDDETLKSKYAALCDVFMVDAEGTAHRAQAASHGIGKPADV
ACAGPLLAELDALGKALKEPARPVAIVGGSKVTKLVLDLSKIAQLIIVGGGIA
NFIQAQHDGKSLYEADLVDEAKRLTLTNPVSDVRVATEFSETAPATLKSYND
NFIQAQHDGKSLYEADLVDEAKRLTLTNPVSDVRVATEFSETAPATLKSYND
FIAGGGDTLAAIDLFIADIKISYISTGGCAFLEFVEGKVLPAVAMLEERAKK"
4401..5480
/note="fructose 1,6-bisphosphate aldolase (AA 1-359)"
/codon_start=1
/transl_table=11
/protein_id="CAA32605.1"
/db_xref="GI:41423"
/db_xref="SWISS-PROT:P11604"
/translacion="MSKIFDVKPGVITGDDVDQKVFQVAKENNEALPAVNCVGTDSIN
AVLETRAQKAVPIVQFNSNGASFTAGKGVKSDVPQGAALIGALSAHVRHOMAEHYG
VPILHTDCAKELLPWIDGLLDAGEKHFATGKPLFSSHMIDLSSELSQENIEICSK
YLERMSKICLAEIEIGTCGDEGDVNSMDASALYTPQEDVDYATYELSKISPRFT
IAASFGNVHGVKPGNVLTPTILRDSQEVYVSKHNLPHNSLNFVFGHGGSGTAQEI
DSVSYGVKRNMDTDTQWATWEGVLNYKANEAYLQGLGNPKGEDQPNKKYDPRVW
LRAGQTSMTARLEKAFQELNAIDL"
5545..5573
/note="repetitive extragenic palindromic sequence"
5580..5586
/note="inverted repeat E"
5597..5603
/note="inverted repeat E"
complement(5613..5762)
/note="repetitive extragenic palindromic sequence"
5838..6698
/note="ORF4 (AA 1-197)"
/codon_start=1
/transl_table=11
/protein_id="CAA32606.1"
/db_xref="GI:41424"
/db_xref="SWISS-PROT:P11666"
/translacion="MEDLNVDINSINGAGSWLVANQALLSYAVNVAALIIIVGLII
ARMINAVNRLMSKIDATVAFSLVRYGIIATFLIAALGRVGVOTASVIAVLGA
AGLVAGLALQSLNSLAAGLVLMRPPFAGEYVDLGGVAGTVLSQVIFSTPMRTADG
KLIIVPNGKIIAGNIINFSRPNREFIIGVAYSDIDQVAKOILNFIQSDRLIKO
REMTVRLNELGASSINFFVVRVWNSGDLQNVYWDVLERIKREFDAAGISFPYPQMDVN
FKRVKEDKAA"
6879..7472
/note="ORF5 (AA 1-197)"
/codon_start=1
/transl_table=11
/protein_id="CAA32607.1"
/db_xref="GI:41425"
/db_xref="SWISS-PROT:P11667"
/translacion="MILPGQNAFVMQIRROVHIMIALCAISDLVICAGIFGG
SALLMQSPMLIALVTGGVAFLLWYFGFAKTMSSNIFLAESAEMKQGRWKIIATML
AYTLNPHVYLDFTFVLGSLGGQDVEPKRFALGTISASFLLWFFGLALLAALAPRL
PFAKQRIINLVGVCWMFIALQLARDGTAHAQALFS"
7565..78029
/note="ORF6 (AA 1-155)"
/codon_start=1
/transl_table=11
/protein_id="CAA32608.1"
/db_xref="GI:581079"
/db_xref="SWISS-PROT:P11668"
/translacion="MKFKVIALAALMGISGMAAQANELPDGHVITSGTASVDAPVDI
ATFLAEVNAKDAATAKQADRAVQYISFLELNQIAKDKISSANLRTQPDYQDQG
KSLIKGYRAVRVETVLRQLDKLSLQALKALGNETRSVSLGVAQPDAYKD"
1984 a 1922 c 2114 g 2009 t
```

```
Query Match 8.5%; Score 60.4; DB 2; Length 8029;
Best Local Similarity 51.1%; Pred. No. 3 8e-07;
Matches 142; Conservative 0; Mismatches 136; Indels 0; Gaps 0;

QY 49 tccacettgtttgcatgctgttcctgcatgacggaacaccataacaggttaagcg 108
DB 7126 TCAATATTACTGCTCATGCTGTTTTAAAAGCGCCAAAACCATACACAGCAAGGCT 7067

QY 109 atgccacccccagcgaataatagcagcagatcgcgcgcgccattggacaaaagatcaagc 158
DB 7066 ACCCCGCCCCAGGTGACAGCGCCAGCACACCGCGGACTGTCATCAATACGCGCTGCCA 7007

QY 169 cccaaggtgcccgcgtgaacaaaaagagcgtcagaataataaacacacagagaagaacgcga 228
DB 7006 CCAAAATCCCGCGCAATCAGGACCAATCGCTATAGCACAAAGTAGGCAATCATYA 6947

QY 229 atgagtccttcgcgttaattcctgtttaaataccacagtcatttcgcggtccgattggac 288
DB 6946 ATGTGTTACTGACGACGTATGCTGCTGATTCATCAGAAAAGCATTTTGTGGACGCGGT 6887

QY 289 agtaaaagactgcccccaaaagcagacacctgtaataaa 326
DB 6886 AGGATCATAGCCGCCCAAGTGCAGAGACCTTGAAGATA 6849

RESULT 14
LOCUS AE000375 10362 bp DNA BCT 01-DEC-2000
DEFINITION Escherichia coli K12 MG1655 section 265 of 400 of the complete
genome.
ACCESSION AE000375 U00096
VERSION AE000375.1 GI:1789282
KEYWORDS Escherichia coli K12.
SOURCE Escherichia coli K12.
ORGANISM Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
REFERENCE 1 (bases 1 to 10362)
AUTHORS Blattner, F.R., Plunkett, G. III, Bloch, C.A., Perna, N.T., Burland, V.,
Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F.,
Gregor, J., Davis, N.W., Kirkpatrick, H.A., Goaden, M.A., Rose, D.J.,
Mau, B. and Shao, Y.
TITLE The complete genome sequence of Escherichia coli K-12
JOURNAL Science 277 (5331), 1453-1474 (1997)
MEDLINE 97426617
PUBMED 9278503
REFERENCE 2 (bases 1 to 10362)
AUTHORS Blattner, F.R.
TITLE Direct Submission
JOURNAL Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics,
University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
Email: ecollegenetics.wisc.edu Phone: 608-262-2534 Fax:
608-263-7459
REFERENCE 3 (bases 1 to 10362)
AUTHORS Blattner, F.R.
TITLE Direct Submission
JOURNAL Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics,
University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
Email: ecollegenetics.wisc.edu Phone: 608-262-2534 Fax:
608-263-7459
REFERENCE 4 (bases 1 to 10362)
AUTHORS Plunkett, G. III.
TITLE Direct Submission
JOURNAL Submitted (13-OCT-1998) Laboratory of Genetics, University of
Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
COMMENT This sequence was determined by the E. coli Genome Project at the
University of Wisconsin-Madison (Frederick R. Blattner, director).
Supported by NIH grants HG00301 and HG01428 (from the Human Genome
Project and NCHGR). The entire sequence was independently
determined from E. coli K12 strain MG1655. Predicted open reading
frames were determined using GeneMark software, kindly supplied by
Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA.
```

30332 [e-mail: mark@amber.gatech.edu]. Open reading frames that have been correlated with genetic loci are being annotated with CG Site Nos., unique ID nos. for the genes in the E. coli Genetic Stock Center (CGSC) database at Yale University, kindly supplied by Mary Berlyn. A public version of the database is accessible (<http://cgsc.biology.yale.edu>). Annotation of the genome is an ongoing task whose goal is to make the genome sequence more useful by correlating it with other data. Comments to the authors are appreciated. Updated information will be available at the E. coli Genome Project's World Wide Web site (<http://www.genetics.wisc.edu>). *** The E. coli K12 sequence and its annotations are periodically updated; this is version M54. No sequence changes. Annotation updates: updated gene identifications and products; all new functional assignments courtesy of Monica Riley; added promoters, protein binding sites, and repeated sequences described in reference 1. The unique numeric identifiers beginning with a lowercase 'b' assigned to each gene (protein- or RNA-encoding) are now designated as gene synonyms instead of labels. This should allow them to be searched for in Entrez as gene names.

FEATURES

```

source      Location/Qualifiers
1. .10362
  /organism="Escherichia coli K12"
  /strain="K12"
  /sub_strain="MG1655"
  /db_xref="taxon:83333"
  complement(9. .40)
  /note="factor Sigma70; predicted +1 start at 3057674"
34. .63
  /note="factor Sigma70; predicted +1 start at 3057742"
101...994
  /gene="icIA"
  /note="b2916"
101...994
  /gene="icIA"
  /function="regulator; DNA - replication, repair,
restriction/modification"
  /note="o297; 100 pct identical to ICIA_ECOLI SW: P24194;
CG Site No. 28646"
  /codon_start=1
  /transl_table=11
  /product="replication initiation inhibitor, binds to
13-mers at oriC"
  /protein_id="AAC75953.1"
  /db_xref="GI:1789283"
  /translation="MKRPDYRTLQALDAVIRERGFERRAOKLCITQSAVORIKQLEN
MFQPLIVRVPPRPVPOGKLLALLRQVLEELWGLDEGTSTPLLSLVANADSL'
ATWLLPALAVLADSPIRLNLQVEDSTRTOERLRGCEVGVCAVSIOHQALPCLVDKLG
ALDYLVSRRPFAEKYFPNGVTRSAKKAPVAFDLDHMQAFLOQNFOLPFGSVPC
HIVNSSEAFVQLARQGTCTCCMIPHLOIEKELASGELIDLPGLFQRRMLYWHRFAPES
RMRKVTDALDLYGHKVLQD"
1045. .1072
  /note="factor Sigma70; predicted +1 start at 3058751"
1198. .3342
  /gene="sbm"
  /note="b2917"
1198. .3342
  /gene="sbm"
  /function="enzyme; Degradation of small molecules: Fatty
acids"
  /note="o714; 99 pct identical to SBM_ECOLI SW: P27253"
  /codon_start=1
  /transl_table=11
  /product="methylmalonyl-CoA mutase (MCM)"
  /protein_id="AAC75954.1"
  /db_xref="GI:1789284"
  /translation="MSNVQEQWLANKELSRREKTVDSLQVHTAEGIAIKPLYTEADL
DNLEVGTLPGLPPYVRGPATMTAQPTIROYAGFSTAKESNAFYRNLAAGQKGL
SVAFDLATHRGSDNPRVAGDVGKAGVADITVEDMKLVFDQIPLDKMSVSMTMNGAV
LPVLAFVIAVEEQVTPDKLTGTIQNDILKLCRTIYIPPKPSMRIADIATJAWS
GNMPRENTISISYHNGEAGACVQVAFTLADGIEYIKAAISAGLKIDDFAPRLSFF
FGIGMDLFNNVLAARVYLWSEAVSGFQAQDPKSLALRTHCOTSGNSLTEQDPYNNV
IRTIIEALATLGTSLHTNADFALGPTDFDSARIARNTQIIIQESELCTVDPL

```

```

AGSYITESLTDQIVKQARATIQIIDEAGGMAKAEAGLPRKRMTEESARQSLDQCK
RVTGVNKKYKLDHEDTDVLEIDNVNVRNEQIASLERIRATRDDAAVTAALNALTHAA
QHNEILAAAVNAARVATRGATGEISDALEAFDRYLVPSQCVTVGIAQSYHQSEKSAFE
FDAIQAQTEFLADNGRRRPRILLIAKMGQDGHGRGAKVIASAYSDLGDFDVLSPMFSTP
EETARLAVENDVHVVGASSIAAGHKTLIPELVEALKKWGREDDICVVGAGVIPPQDYAF
LQERGVAALYGPCTPMLDSVRDVLNLSQHH"
3335. .4330
  /gene="ygfD"
  /note="b2918"
3335. .4330
  /gene="ygfD"
  /function="phenotype; Not classified"
  /note="o331; 100 pct identical to YGF_D_ECOLI SW:
P27254(257 aa) but has 25 additional N-ter aa 50
additional C-ter aa; orf2 of X66836"
  /codon_start=1
  /transl_table=11
  /product="putative nucleotide-binding protein"
  /protein_id="AAC75955.1"
  /db_xref="GI:1789285"
  /translation="MINEATLAESIRRLROGERATLAQAMTLVESRHRHQALSTQLL
DAIMPYCGNTLRGVCTPGAGKSTFEAFGMLLIREGLKVAIVADVPSPVTVGGSTIL
GDKTRMNDLARAEEAATRPVPSGHLGGASORARELMLLCEAAGYDVIVIVETVGVQGS
ETEVARNVDCFISLQIAGGDDLOGIKKGLMEVADLIVINKDDGDNHTNVAIARHMYE
SALHILRRKYDQWQRPVLTCSALEKRGIDELIWIHAIIDFKTALTASRLGRLOOVRQOQSV
WLRKQTEEEVLNHLFANEDFDYRYRQTLAIVKNTLSPRTGLRQLSEFIQTQYFD"
4299. .5126
  /gene="ygfG"
  /note="b2919"
4299. .5126
  /gene="ygfG"
  /function="putative enzyme; Not classified"
  /note="o275; This 275 aa ORF is 30 pct identical (6 gaps)
to 250 residues of an approx. 296 aa protein ECHM_RAT SW:
P14604"
  /codon_start=1
  /transl_table=11
  /product="putative enzyme"
  /protein_id="AAC75956.1"
  /db_xref="GI:1789286"
  /translation="MNLSSRNLIKIFMSYQVNVVINKVAVIEFNFGKRLNALSK
VFDDLMQALSDLNRPETICILRAPSGSKVFSAGHDHIEPLSGRDLPSYDDPLRQI
TRMOKFPRPIISWEGSVWGAFEMMSDILLIAASTFTSMTPNVNLGVNPLVCGIHL
NLTRDAGFHIVKELIIFASPTAQRALAVGLNHWVEVELEDFTLQMAHHISEKAPL
AIVAIKEULRVGLGEAHTMSDEFEIQGMRRAVYDSEYQEGMNAFLKKRPVFGH"
5150. .6628
  /gene="ygfH"
  /note="b2920"
5150. .6628
  /gene="ygfH"
  /function="putative enzyme; Not classified"
  /note="o492; This 492 aa ORF is 46 pct identical (7 gaps)
to 490 residues of an approx. 544 aa protein CAT1_CLOCK
SW: P38946"
  /codon_start=1
  /transl_table=11
  /product="putative coenzyme A transferase"
  /protein_id="AAC75957.1"
  /db_xref="GI:1789287"
  /translation="METQWTRMTANAEALIOHNDMVAFSGFTPAGSKALPTAIARR
ANQHEAKRPYQIRLLTGASISAAADDVLSDADAVSWRAPYQTSGLRKKINGAVSF
VDHLSEVAQMVNYPFGDIDDAVIEASALAPDGRVWLTSIGNAPTWLLRAKKVIIIE
LNHYHDPVRAELADIVIPGAPPRNSVIFHAMDRVGTQVQIDPKKIVAVVETNLPD
AGNMDKQNPQCOQIADNVVTFLQEMAHGRIPPEELPQSGVGNKNNVAVARLGENP
VIPPFMVSEVLOQSVHLLLETKGISASASLITISADSLRKIYDNMDYFASRIVLRP
QELSNPEIIRRLGVALANVLEFDIYGHANSTHVAGVDLMNGICSGSGDFERNAYLSI
FMAPSIAKEKISTVYPMCSHVHDSEHSKVITITEOGIADLRGLSLPQARATIIDNCA
HPMYRDYLKRIENAPGHHHDLSHVFDLHRNLIAATGMLG"
  complement(6625. .7536)
  /gene="ygfI"
  /note="b2921"
  complement(6625. .7536)
  /gene="ygfI"

```



```
misc_feature
GETAIVQIRNREMPVKVTKPVVRNGKAVA"
/complement(5370..10317)
/notes="similar to GenBank Accession Number D90281
(ECOPUV)"
gene
complement(5506..6708)
/gene="visc"
CDS
complement(5506..6708)
/gene="visc"
/notes="ORF_f400"
/codon_start=1
/transl_table=11
/protein_id="AAA69074.1"
/db_xref="GI:882436"
/transl_table="MOSVDVAIVGGMVLAVACGGGSLRVAVLEQRVQVEPLAANA
PPQLRVSAINAASEKLLRLGVDLIRASCYHGMVWMDKDSFGHISFDDQSMGYS
HLGHIVENSVIHYALNKAHQSSDITLLAPAEQVQAWGENETFLTLDGCSMLTARLV
IGADGANSWLRNADIPLTFWDYOHHALVATIRTEPHDAVARQVFGHIGLAFPLPS
DPLHCSIVWSLSPPEAQMOQASEDENRALNIAFONRLGLCKVESAROVFPITGRVA
RPFASHELALVGDAHTIHPLAGOGYNLGPMDAAELIAELKRLHROCKDQGVLYLRR
YERSKRHSALMAGMQGFRDLFSGYNPAKKLLRDIGKLADTLPGVKPOLIRQAMGL
NDLPEWLR"
/complement(6731..7909)
/gene="ubih"
gene
complement(6731..7909)
CDS
/gene="ubih"
/notes="ORF_f392"
/codon_start=1
/transl_table=11
/protein_id="AAA69075.1"
/db_xref="GI:882437"
/transl_table="MSVIVGGMGATLALISRLSHGALPVHLTEATPESHAPHG
FDGRALAAAGTQQQLARQVWQSLADCATATTVHSDRGHAGFVTLAAEDYOLAAAL
GOVEHLNVGQRFALLRKAPGVTLHCPDRVANVARTOSHEVTLSEGETLGRVLVA
ADGTHSALATACQVDWQOEPEYEQLVIANVATVAHEGRAFERFTQHGPLMLPMSDG
RCSLVCHPEREREWSDEKFCRELOSAFGWRGLKIITHAGKRAYPLALTHAARS
IHTWLVGNAAOTLHPLAGOGNLCMRDWSLAETLTOAQERGEDMGDGVLCRVQO
RROSDREATIGVDSLVHLFANRWAPLVVGRNIGLMTWELFTFARDVLAORTLGWVAR"
/complement(7749..9599)
/notes="similar to GenBank Accession Number D00398
(ECOAP2)"
gene
complement(7906..9231)
CDS
/gene="pepp"
complement(7906..9231)
/notes="ORF_f441; third start codon"
/codon_start=1
/transl_table=11
/protein_id="AAA69076.1"
/db_xref="GI:882438"
/transl_table="MSEISRQFRRRQALVEQMPGSAALIFAAPEVTRSADSEYPY
RQNSDFWYFTGFNEPVALVLIKSDDTNHNHSLVFNVRDLTAEIFGRRLGQDAAPEK
LGVDRALAFSEIQOLYOLLGDDVYVHAQGEYADYIVNSALEKLRKGSRONLTAP
ATMIDRPPVHMRLEKSPPEIAYLRAGEITAMATRAKEKCRPGMEFYHLEGEIHH
FNRHGARYSPYTIIVGSGGICLHYTENECERDGLVLDIADGCEIKGYAGDITRT
FPVNGKFTQRYEIVDLSLETSLRLXRPGTSLIEVTGEVVRIMVSGLVKLGILKG
DVEDLIAQNAHRPFMHGLGDLVDHGVYCDRSRILEPCGMVLTPEPGLYIAPD
AEVPEQYRGIGIEDDITVITETGNNENLTASVVKPPEIEALMVAARKQ"
/gene="ygfB"
complement(9257..9841)
CDS
/gene="ygfB"
/notes="ORF_f194"
/codon_start=1
/transl_table=11
/protein_id="AAA69077.1"
/db_xref="GI:882439"
/transl_table="MLMSIQNEMPGYNEMNQYNQOQGTGLTPAEMHGLISGMICGGND
DSSWLPLLHDLTNEGMAFGHELAAOLARKMHSATSDALODDGFULYLPDGDVSVFD
RADLAGWVNHLLGLGVTPQKLDKVTGETGEAIDDLRNIAQLGYDEDEDEQELENSL
EEIIEYRVAAALICDHTFTHPQTAPEVOKPTUH"
/10003..10332
/notes="ORF_o109"
```

```
misc_feature
10266..11219
/notes="similar to GenBank Accession Number M12965
(ECOSSR)"
misc_RNA
10374..10556
/gene="ssr"
/product="stable 6S RNA"
gene
10374..10556
/gene="ssr"
misc_feature
10374..10557
/notes="similar to GenBank Accession Number X01238
(ECRRN6S)"

Query Match      8.58; Score 60.4; DB 2; Length 141744;
Best Local Similarity 51.18; Pred. No. 7.1e-07;
Matches 142; Conservative 0; Mismatches 136; Indels 0; Gaps 0;

Qy 49 tccacctgtttgtcattggtcttcttcgctgcatgacgcaaacaccataacagtaacgcg 108
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 22910 TCAATATTACTGCTCATTTGCTGTTTTAAAGCGCCAAACCATACACAGCAAGAGGCT 22969
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 109 atgccacccagcgcaataatcagcagcagtcgcgcgcgccattggacaaaagatcaacg 168
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 22970 ACGCCGCCCGAGTGACCGCGCCAGCAACCCAGCGGACTGCATCAATAAGCGCTGCCA 23029

Qy 169 cccaaagtgcggcgatgaacaaagacgctcagaaattaaacacacagagaagaccgca 228
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 23030 CCAAAAATCCCGCGCAATATCAGGACCAAAATCGCTGATAGCACAAAGTAAGGCAATCATA 23089

Qy 229 atgaatccttcgagcttaattccttctgtttaaaccacagatcattcgcgtccgatggac 288
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 23090 ATGTGGTACTGACGACGATGTCCTGATTCATCACAAAAGCATTTTGTGGACCGCGGT 23149

Qy 289 agtaaaagactggcccccaaaagcagacctgtaatgaa 326
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 23150 AGGATCATAGCCGCCCAAGTGCAAGACCTTGAAAGTA 23187

Search completed: September 14, 2001, 07:52:13
Job time: 15035 sec
```


Result No.	Score	Query		DB	ID	Description
		Match	Length			
1	63.2	8.9	15239	1	US-08-390-878-17	Sequence 17, Appl
C 2	37	5.2	2058	1	US-08-358-117-1	Sequence 1, Appli
C 3	37	5.2	2058	3	US-08-470-588-1	Sequence 1, Appli
4	33	4.6	2656	2	US-08-683-625A-5	Sequence 5, Appli
5	32.8	4.6	1876	3	US-08-714-918-33	Sequence 33, Appl
6	32.8	4.6	1876	4	US-09-265-315-33	Sequence 33, Appl
7	32.8	4.6	1876	4	US-09-265-315-33	Sequence 33, Appl
8	32.8	4.6	1876	4	US-09-266-417-33	Sequence 33, Appl
9	31.2	4.4	219	3	US-08-078-271B-9	Sequence 9, Appli
10	31	4.4	4700	4	US-09-150-460B-9	Sequence 9, Appli
C 11	31	4.4	7218	1	US-08-232-463-14	Sequence 14, Appl
C 12	30.8	4.3	8501	3	US-08-793-900-1	Sequence 1, Appli
C 13	30.6	4.3	1299	4	US-09-385-028-1	Sequence 21, Appl
C 14	30.6	4.3	11604	4	US-09-385-028-13	Sequence 13, Appl
C 15	30.6	4.3	13079	4	US-09-385-028-1	Sequence 1, Appli
C 16	30.2	4.2	9207	3	US-08-388-353-800	Sequence 800, Appl
C 17	30.2	4.2	9207	3	US-08-488-531B-800	Sequence 800, Appl
C 18	30	4.2	213	1	US-07-867-194-3	Sequence 3, Appli
C 19	30	4.2	259	1	US-07-867-194-4	Sequence 4, Appli
C 20	30	4.2	261	1	US-07-867-194-5	Sequence 5, Appli
C 21	29.6	4.2	1872	1	US-08-428-926-1	Sequence 1, Appli
C 22	29.6	4.2	1872	1	US-08-435-434-4	Sequence 4, Appli
C 23	29.6	4.2	1872	1	US-08-435-436-4	Sequence 4, Appli
C 24	29.6	4.2	1872	1	US-08-428-927-1	Sequence 1, Appli
C 25	29.6	4.2	1872	1	US-08-428-298-1	Sequence 1, Appli
C 26	29.6	4.2	1872	1	US-08-339-517-1	Sequence 1, Appli
C 27	29.6	4.2	1872	2	US-08-438-863-4	Sequence 4, Appli


```

; GENERAL INFORMATION:
; APPLICANT: UENO, Naoto
; APPLICANT: MATSUMOTO, Kunihiko
; APPLICANT: IRIE, Kenji
; TITLE OF INVENTION: NOVEL KINASE IN TGF-BETA FAMILY SIGNAL
; TITLE OF INVENTION: TRANSDUCTION SYSTEM
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, L.L.P.
; STREET: P.O. Box 1404
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/685,625A
; FILING DATE: 24-JUL-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 7-253549
; FILING DATE: 29-SEP-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Meuth, Donna M.
; REGISTRATION NUMBER: 36,607
; REFERENCE/DOCKET NUMBER: 001560-267
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2656 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 183..1922
; US-08-685-625A-5

Query Match 4.6% Score 33; DB 2; Length 2656;
Best Local Similarity 63.0% Pred. No. 0.79;
Matches 51; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

Qy 499 gtgagtcagcgcgttaagctctcgagcatcacgtgggtcagtggtgtatcgcgacc 558
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1264 GTGAATCTGGACGTTTAACTTGGAGCCTCCCATGGGAGCAGTCTGGAGAGCTGCCCC 1323

Qy 559 caaccggccaaagcaaccgaa 579
||||| | ||| | |
Db 1324 CAACCTCTGAGGCAAGAGGA 1344

RESULT 5
US-08-714-918-33
; Sequence 33, Application US/08714918
; Patent No. 6037123
; GENERAL INFORMATION:
; APPLICANT: Benton, Bret
; APPLICANT: Lee, Ving
; APPLICANT: Malouin, Francois
; APPLICANT: Martin, Patrick K.
; APPLICANT: Schmid, Molly B.
; APPLICANT: Sun, Dongxu
; TITLE OF INVENTION: STAPHYLOCOCCUS AUREUS ANTIBACTERIAL
; TITLE OF INVENTION: TARGET GENES
; NUMBER OF SEQUENCES: 111

```

```

; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Fifth Street
; STREET: Suite 4700
; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90071-2066
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; MEDIUM TYPE: storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: IBM P.C. DOS 5.0
; SOFTWARE: Word Perfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/714,918
; FILING DATE: September 13, 1996
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/009,102
; FILING DATE: December 22, 1995
; APPLICATION NUMBER: 60/003,798
; FILING DATE: September 15, 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Warburg, Richard J.
; REGISTRATION NUMBER: 32,327
; REFERENCE/DOCKET NUMBER: 222/005
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 33:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1876 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-714-918-33

Query Match 4.6% Score 32.8; DB 3; Length 1876;
Best Local Similarity 48.3% Pred. No. 0.77;
Matches 70; Conservative 7; Mismatches 68; Indels 0; Gaps 0;

Qy 292 aaagactggcccccacaaagcagacctgtaagaagatttccatgatccaccatcgtagcc 351
||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 606 ARAHTAATTATCTTCADCCAHAGAAAGGATTGATGATTTCGTGTCMTCAATTAGAAG 665

Qy 352 tatggaagtcacttaagtaaatgattggttcttaacatggtttaatatagcttcataaac 411
|||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 666 AATGGTTTGAGATATKCGACAGTTGGTTTWTACCTGGATTATATTGTTATATATTA 725

Qy 412 cccattcaactggacacttgctct 436
|||| | | | | | | | | | |
Db 726 GAGCTNTAATTCAGTATTTCCCTTT 750

RESULT 6
US-09-265-315-33
; Sequence 33, Application US/09265315
; Patent No. 6187541
; GENERAL INFORMATION:
; APPLICANT: Benton, Bret
; APPLICANT: Lee, Ving J.
; APPLICANT: Malouin, Francois
; APPLICANT: Martin, Patrick K.
; APPLICANT: Schmid, Molly B.
; APPLICANT: Sun, Dongxu
; TITLE OF INVENTION: METHODS OF SCREENING FOR COMPOUNDS
; TITLE OF INVENTION: ACTIVE ON STAPHYLOCOCCUS AUREUS
; TITLE OF INVENTION: TARGET GENES
; NUMBER OF SEQUENCES: 111
; CORRESPONDENCE ADDRESS:

```

ADDRESSEE: Lyon & Lyon
STREET: 633 West Fifth Street
STREET: Suite 4700
CITY: Los Angeles
STATE: California
COUNTRY: U.S.A.
ZIP: 90071-2066
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
MEDIUM TYPE: storage
COMPUTER: IBM Compatible
OPERATING SYSTEM: IBM P.C. DOS 5.0
SOFTWARE: Word Perfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/265,315
FILING DATE: March 9, 1999
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/714,918
FILING DATE: September 13, 1996
APPLICATION NUMBER: 60/009,102
FILING DATE: December 22, 1995
APPLICATION NUMBER: 60/003,798
FILING DATE: September 15, 1995
ATTORNEY/AGENT INFORMATION:
NAME: Warburg, Richard J.
REGISTRATION NUMBER: 32,327
REFERENCE/DOCKET NUMBER: 240/247
TELEPHONE: (213) 489-1600
TELEFAX: (213) 955-0440
TELEX: 67-3510
INFORMATION FOR SEQ ID NO: 33:
SEQUENCE CHARACTERISTICS:
LENGTH: 1876 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-265-315-33

Query Match 4.6%; Score 32.8; DB 4; Length 1876;
Best Local Similarity 48.3%; Pred. No. 0.77; Indels 0; Gaps 0;
Matches 70; Conservative 7; Mismatches 68; Indels 0; Gaps 0;
QY 292 aaagactgcccccaaaagcagacctgaatgaagatttccatgatccaccatcgtgacc 351
Db 606 ARAHTAATTATCTTCADCCAHAGAAAGGATTCGATGTTGCGTWTCTCAATTAGAAG 665
QY 352 tatggaagtacttaagaaatgattggttcttaacatggttttaataatagcttcataaac 411
Db 666 AATGGTTTGAGATATKTCGACAGTTGGTTTWTACCTGGATTATATGTTATATATTA 725
QY 412 cccattcaactggacactttgctct 436
Db 726 GAGCTNTAATCCAGTATTCCTTT 750

RESULT 7
US-09-265-315-33
Sequence 33, Application US/09265315
Patent No. 6187541
GENERAL INFORMATION:
APPLICANT: Benton, Bret
APPLICANT: Lee, Ving J.
APPLICANT: Malouin, Francois
APPLICANT: Martin, Patrick K.
APPLICANT: Schmid, Molly B.
APPLICANT: Sun, Dongxu
TITLE OF INVENTION: METHODS OF SCREENING FOR COMPOUNDS
TITLE OF INVENTION: ACTIVE ON STAPHYLOCOCCUS AUREUS
NUMBER OF SEQUENCES: 111

CORRESPONDENCE ADDRESS:
ADDRESSEE: Lyon & Lyon
STREET: 633 West Fifth Street
STREET: Suite 4700
CITY: Los Angeles
STATE: California
COUNTRY: U.S.A.
ZIP: 90071-2066
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
MEDIUM TYPE: storage
COMPUTER: IBM Compatible
OPERATING SYSTEM: IBM P.C. DOS 5.0
SOFTWARE: Word Perfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/265,315
FILING DATE: March 9, 1999
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/714,918
FILING DATE: September 13, 1996
APPLICATION NUMBER: 60/009,102
FILING DATE: December 22, 1995
APPLICATION NUMBER: 60/003,798
FILING DATE: September 15, 1995
ATTORNEY/AGENT INFORMATION:
NAME: Warburg, Richard J.
REGISTRATION NUMBER: 32,327
REFERENCE/DOCKET NUMBER: 240/247
TELEPHONE: (213) 489-1600
TELEFAX: (213) 955-0440
TELEX: 67-3510
INFORMATION FOR SEQ ID NO: 33:
SEQUENCE CHARACTERISTICS:
LENGTH: 1876 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-265-315-33

Query Match 4.6%; Score 32.8; DB 4; Length 1876;
Best Local Similarity 48.3%; Pred. No. 0.77; Indels 0; Gaps 0;
Matches 70; Conservative 7; Mismatches 68; Indels 0; Gaps 0;
QY 292 aaagactgcccccaaaagcagacctgaatgaagatttccatgatccaccatcgtgacc 351
Db 606 ARAHTAATTATCTTCADCCAHAGAAAGGATTCGATGTTGCGTWTCTCAATTAGAAG 665
QY 352 tatggaagtacttaagaaatgattggttcttaacatggttttaataatagcttcataaac 411
Db 666 AATGGTTTGAGATATKTCGACAGTTGGTTTWTACCTGGATTATATGTTATATATTA 725
QY 412 cccattcaactggacactttgctct 436
Db 726 GAGCTNTAATCCAGTATTCCTTT 750

RESULT 8
US-09-266-417-33
Sequence 33, Application US/09266417
Patent No. 6228588
GENERAL INFORMATION:
APPLICANT: Benton, Bret
APPLICANT: Lee, Ving J.
APPLICANT: Malouin, Francois
APPLICANT: Martin, Patrick K.
APPLICANT: Schmid, Molly B.
APPLICANT: Sun, Dongxu
TITLE OF INVENTION: METHODS OF SCREENING FOR COMPOUNDS
TITLE OF INVENTION: ACTIVE ON STAPHYLOCOCCUS AUREUS
TITLE OF INVENTION: TARGET GENES

```

: NUMBER OF SEQUENCES: 111
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Lyon & Lyon
: STREET: 633 West Fifth Street
: STREET: Suite 4700
: CITY: Los Angeles
: STATE: California
: COUNTRY: U.S.A.
: ZIP: 90071-2066
: COMPUTER READABLE FORM:
: MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
: MEDIUM TYPE: storage
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: IBM P.C. DOS 5.0
: SOFTWARE: Word Perfect 5.1
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/266,417
: FILING DATE: March 9, 1999
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/714,918
: FILING DATE: September 13, 1996
: APPLICATION NUMBER: 60/009,102
: FILING DATE: December 22, 1995
: APPLICATION NUMBER: 60/003,798
: FILING DATE: September 15, 1995
: ATTORNEY/AGENT INFORMATION:
: NAME: Warburg, Richard J.
: REGISTRATION NUMBER: 32,327
: REFERENCE/DOCKET NUMBER: 240/248
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (213) 489-1600
: TELEFAX: (213) 955-0440
: TELEX: 67-3510
: INFORMATION FOR SEQ ID NO: 33:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1876 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: US-09-266-417-33

```

	Query Match	4.6%	Score 32.8	DB 4	Length 1876
	Best Local Similarity	48.3%	Pred. No. 0.77		
	Matches 70	Conservative 7	Mismatches 68	Indels 0	Gaps
Qy	292	aaagactggccccc	aaagcagacacgtgaatgaagatttccatgatcaccatcgtgacc	351	
		:	:		
Db	606	ARHTTAATATCTC	ADCCAHAGAAAGGGATTGATGATTTCGTWTCMTCAATTAGAAG	665	
Qy	352	tatggaagtactlaag	tcaaatgatitggttcttcaacatggtttcaatgatcctcatgaac	411	
Db	666	AATGGTTTGAGATAT	TCGCACAGTGGTGGTWTWTTCCTGGATTATATGTTATATATTA	725	
Qy	412	cccattcaactggac	actttgctct	436	
Db	726	GAGCTNTAATTC	CAGTATATTCCTTT	750	

RESULT 9
US-08-078-271B-9
; Sequence 9, Application us/08078271B
; Patent No. 6136527
; GENERAL INFORMATION:
; APPLICANT: Fuchs, Klaus; Motz, Manfred; Roggendorf, Michael;
; APPLICANT: Soutschek, Erwin
; TITLE OF INVENTION: Polypeptides Derived from Proteins of the
; TITLE OF INVENTION: Hepatitis C Virus, Test Kits Containing these Polypeptides
; TITLE OF INVENTION: Hepatitis C Virus and Vaccines against Infection by Hepatitis C Virus
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felfe & Lvnch

STREET: 805 Third Avenue
CITY: New York City
STATE: New York
COUNTRY: USA
ZIP: 10022

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
COMPUTER: IBM PS/2
OPERATING SYSTEM: PC-DOS
SOFTWARE: Wordperfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/078, 271B
FILING DATE: 15-OCT-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/DE91/01020
FILING DATE: 20-DEC-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: DE 40 41 304.7
FILING DATE: 21-DEC-1990
ATTORNEY/AGENT INFORMATION:
NAME: Hanson, No. 6136527man D.
REGISTRATION NUMBER: 30,946
REFERENCE/DOCKET NUMBER: LEADER 201
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 688-9200
TELEFAX: (212) 838-3884
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 219 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear

US-08-078-271B-9

Query Match 4.4%; Score 31.2; DB-3; Length 219;
Best Local Similarity 50.7%; Pred. No. 0.85;
Matches 75; Conservative 0; Mismatches 73; Indels

Qy	105	agcgatgcccaccccgacgcaataatcagagcacagatcgcgcgcgccattgagacaaaagatc	164
Db	31	AGGGGGGCTGGCGAGGGGACGGCATATGCACACTCGCTGCCACTGTGGAGCTGAGATC	90
Qy	165	aagcgccaaagtgccgcgcgatgacaaaaagacgctcagaataataaacacacagagaagac	224
Db	91	ACCGGACATCTCAAGAACGGGACGATGAGGATCTCGGTCTTAAGACCTGCAGGAACATG	150
Qy	225	cgcgaatgagtccttcgcgttaattcct	252
Db	151	TGAGAGTGGGACCTTCGCCATTATGCT	178

```

RESULT 10
US-09-150-460B-9
; Sequence 9, Application US/09150460B
; Patent NO. 6190882
; GENERAL INFORMATION:
; APPLICANT: Lee, Cheng-Chi
; APPLICANT: Albrecht, Urs
; APPLICANT: Eichelle, Gregor
; APPLICANT: Sun, Zhong Sheng
; TITLE OF INVENTION: Mammalian Circadian Rhythm-Like Gene
; FILE REFERENCE: D6039
; CURRENT APPLICATION NUMBER: US/09/150,460B
; CURRENT FILING DATE: 1998-09-09
; PRIOR APPLICATION NUMBER: US 60/058,256
; PRIOR FILING DATE: 1997-09-09
; NUMBER OF SEQ ID NOS: 21
; SEQ ID NO 9
; LENGTH: 4700
; TYPE: DNA
; ORGANISM: artificial sequence

```


	Query Match	4.3%	Score 30.6;	DB 4;	Length 11604;
	Best Local Similarity	53.8%	Pred.No. 10;		
	Matches 63;	Conservative	0;	Mismatches 54;	Indels 0; Gaps 0;
QY	449	aagcgagcttcgaagcgccctcttagcccttccattccccctggcggtgagtcagc	508		
Db	9249	ACGGCAGCATCAACCGGGCGCGGGTGTGTGATCAGCCAGTCGCGCTCACCGCC	9190		
QY	509	gcgttaaagctctcagacatcacgtgggtcgagtggttggtatcgcgaccacacgg	565		

RESULT 15

	Query Match	4.3%	Score 30.6;	DB 4;	Length 15079;
	Best Local Similarity	53.8%;	Pred. No. 11;		
	Matches 63;	Conservative 0;	Mismatches 54;	Indels 0;	Gaps 0;
QY	449	aaggcagcttcgaaggcgctcctcttagccctttccattcccccctcgccgctgagtcagc	508		
Db	11281	ACGGCAGCATCAACCGGCGCGCGGTGTGTGATCAGCCAGTCGCGGTCAACGCC	11222		
QY	509	gccttaagctctcgagcatcagtggtgcagtggtgtggtatcgcgacccacacccgg	565		
Db	11221	GGATTCAACCGGCTGGAGACATCCCTCGGCGCCGGCTGCTGCTCGGACGCCGGGG	11165		

Search completed: September 14, 2001, 07:51:13
Job time: 12387 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 14, 2001, 03:38:13 ; Search time 1873.17 seconds
(without alignments)
3588.024 Million cell updates/sec

Title: US-09-105-117i-1_COPY_1016_1726
Perfect score: 711
Sequence: 1 ggtcgtcgggcacggttg.....cagattcgctacacatgg 711

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 10228115 seqs, 4726426750 residues

Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST.*
1: gb_est1.*
2: gb_est2.*
3: gb_est3.*
4: gb_est4.*
5: gb_est5.*
6: gb_est6.*
7: gb_est7.*
8: gb_est8.*
9: gb_est9.*
10: gb_est10.*
11: gb_est11.*
12: gb_est12.*
13: gb_est13.*
14: gb_est14.*
15: gb_est15.*
16: gb_est16.*
17: gb_est17.*
18: gb_est18.*
19: gb_est19.*
20: gb_est20.*
21: gb_est21.*
22: gb_est22.*
23: gb_est23.*
24: gb_est24.*
25: gb_est25.*
26: gb_est26.*
27: gb_est27.*
28: gb_est28.*
29: gb_est29.*
30: gb_est30.*
31: gb_est31.*
32: gb_est32.*
33: gb_est33.*
34: gb_est34.*
35: gb_est35.*
36: gb_est36.*
37: gb_est37.*
38: gb_est38.*
39: gb_est39.*
40: gb_est40.*
41: gb_est41.*
42: gb_est42.*
43: gb_est43.*
44: gb_est44.*
45: gb_est45.*
46: gb_est46.*
47: gb_est47.*

44: em_esthum10.*
45: em_esthum11.*
46: em_esthum12.*
47: em_esthum13.*
48: em_esthum14.*
49: em_esthum15.*
50: em_esthum16.*
51: em_esthum17.*
52: em_esthum18.*
53: em_esthum19.*
54: em_esthum20.*
55: em_esthum21.*
56: em_esthum22.*
57: em_esthum23.*
58: em_esthum24.*
59: em_esthum25.*
60: em_esthum26.*
61: em_esthum27.*
62: em_esthum28.*
63: em_estin1.*
64: em_estin2.*
65: em_estin3.*
66: em_estin4.*
67: em_estin5.*
68: em_estom1.*
69: em_estom2.*
70: em_estov1.*
71: em_estov2.*
72: em_estpl1.*
73: em_estpl2.*
74: em_estpl3.*
75: em_estpl4.*
76: em_estpl5.*
77: em_estpl6.*
78: em_estpl7.*
79: em_estpl8.*
80: em_estpl9.*
81: em_estpl10.*
82: em_estro1.*
83: em_estro2.*
84: em_estro3.*
85: em_estro4.*
86: em_estro5.*
87: em_estro6.*
88: em_estro7.*
89: em_estro8.*
90: em_estro9.*
91: em_estro10.*
92: em_estro11.*
93: em_estro12.*
94: em_estro13.*
95: em_estro14.*
96: em_estro15.*
97: em_estro16.*
98: em_estro17.*
99: em_estro18.*
100: em_estro19.*
101: em_estro20.*
102: gb_est25.*
103: gb_est26.*
104: gb_est27.*
105: gb_est28.*
106: gb_est29.*
107: gb_est30.*
108: gb_est31.*
109: gb_est32.*
110: gb_est41.*
111: gb_est42.*
112: gb_est43.*
113: gb_est44.*
114: gb_est45.*
115: gb_est46.*
116: gb_est47.*

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est67:*
137: gb_est68:*
138: gb_est69:*
139: gb_est70:*
140: gb_est71:*
141: gb_est72:*
142: gb_est73:*
143: gb_est74:*
144: gb_est75:*
145: gb_est76:*
146: gb_est77:*
147: gb_est78:*
148: gb_est79:*
149: gb_est80:*
150: gb_est81:*
151: gb_est82:*
152: gb_est83:*
153: gb_est84:*
154: gb_est85:*
155: gb_est86:*
156: gb_est87:*
157: gb_est88:*
158: gb_est89:*
159: gb_est90:*
160: gb_est91:*
161: gb_est92:*
162: gb_est93:*
163: gb_est94:*
164: gb_est95:*
165: gb_est96:*
166: gb_est97:*
167: gb_est98:*
168: gb_est99:*
169: gb_est100:*
170: gb_est101:*
171: gb_est102:*
172: gb_est103:*
173: gb_est104:*
174: gb_est105:*
175: gb_est106:*
176: gb_est107:*
177: gb_est108:*
178: gb_est109:*
179: gb_est110:*
180: gb_est111:*
181: gb_est112:*
182: gb_est113:*
183: gb_est114:*
184: gb_est115:*
185: gb_est116:*
186: gb_est117:*
187: gb_est118:*
188: gb_est119:*
189: gb_est120:*

190: gb_est121:*
191: gb_est122:*
192: gb_est123:*
193: gb_est124:*
194: gb_est125:*
195: gb_est126:*
196: gb_est127:*
197: gb_est128:*
198: gb_est129:*
199: gb_est130:*
200: gb_est131:*
201: gb_est132:*
202: gb_est133:*
203: gb_est134:*
204: gb_est135:*
205: gb_est136:*
206: gb_est137:*
207: gb_est138:*
208: gb_est139:*
209: gb_est140:*
210: gb_est141:*
211: gb_est142:*
212: gb_est143:*
213: gb_est144:*
214: gb_est145:*
215: gb_est146:*
216: gb_est147:*
217: gb_est148:*
218: gb_est149:*
219: gb_est150:*
220: gb_est151:*
221: gb_est152:*
222: gb_est153:*
223: gb_est154:*
224: gb_est155:*
225: gb_est156:*
226: gb_est157:*
227: gb_est158:*
228: gb_est159:*
229: gb_est160:*
230: gb_est161:*
231: gb_est162:*
232: gb_est163:*
233: gb_est164:*
234: gb_est165:*
235: gb_est166:*
236: gb_est167:*
237: gb_est168:*
238: gb_est169:*
239: gb_est170:*
240: gb_est171:*
241: gb_est172:*
242: gb_est173:*
243: gb_est174:*
244: gb_est175:*
245: gb_est176:*
246: gb_est177:*
247: gb_est178:*
248: gb_est179:*
249: gb_est180:*
250: gb_est181:*
251: gb_est182:*
252: gb_est183:*
253: gb_est184:*
254: gb_est185:*
255: gb_est186:*
256: gb_est187:*
257: gb_est188:*
258: gb_est189:*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	60.4	8.5	830	219	CNS01MQH
c	40	5.6	916	219	CNS015QU
3	37.2	5.2	2084	144	BF128512
c	37	5.2	911	220	CNS031CW
5	37	5.2	1018	152	BG342366
6	36.4	5.1	409	15	AI020087
7	36.4	5.1	473	17	AI174155
c	36.2	5.1	1100	219	CNS014RI
c	36	5.1	436	228	AQ0405681
10	35.8	5.0	650	149	BF485823
c	35.6	5.0	166	115	AW351850
12	35.6	5.0	222	147	BF352878
13	35.6	5.0	393	234	AQ851405
c	35.6	5.0	492	111	AW067177
15	35.6	5.0	494	240	AZ253065
16	35.6	5.0	552	233	AQ782619
17	35.4	5.0	667	236	AQ991044
18	35.4	5.0	819	146	BF269985
19	35.2	5.0	311	250	A2834870
20	35.2	5.0	590	250	A2834861
c	35	4.9	730	32	AV706549
22	34.8	4.9	295	161	BB565230
23	34.8	4.9	404	13	AA879865
24	34.8	4.9	459	12	AA794877
25	34.8	4.9	459	18	AI315494
26	34.8	4.9	479	8	AA497438
27	34.8	4.9	483	165	BE290613
28	34.8	4.9	499	1	AA065568
c	34.8	4.9	546	258	TA193A020
30	34.8	4.9	547	143	BF016664
31	34.8	4.9	552	166	BE303625
32	34.8	4.9	553	114	AW318849
33	34.8	4.9	585	166	BE371127
34	34.8	4.9	590	165	BE282182
35	34.8	4.9	590	168	BF719586
36	34.8	4.9	607	166	BE334491
37	34.8	4.9	692	166	BE377573
38	34.8	4.9	712	16	AI119048
39	34.8	4.9	712	235	AQ941109
40	34.8	4.9	860	145	BF167006
41	34.8	4.9	888	190	WR2115
42	34.8	4.9	926	145	BF140267
43	34.8	4.9	929	174	BF174120
44	34.8	4.9	936	169	BF788835
45	34.8	4.9	1301	192	AK014354

ALIGNMENTS

Result No.	Score	Query Match	Length	ID	Description
1	60.4	8.5	830	219	CNS01MQH
c	40	5.6	916	219	CNS015QU
3	37.2	5.2	2084	144	BF128512
c	37	5.2	911	220	CNS031CW
5	37	5.2	1018	152	BG342366
6	36.4	5.1	409	15	AI020087
7	36.4	5.1	473	17	AI174155
c	36.2	5.1	1100	219	CNS014RI
c	36	5.1	436	228	AQ0405681
10	35.8	5.0	650	149	BF485823
c	35.6	5.0	166	115	AW351850
12	35.6	5.0	222	147	BF352878
13	35.6	5.0	393	234	AQ851405
c	35.6	5.0	492	111	AW067177
15	35.6	5.0	494	240	AZ253065
16	35.6	5.0	552	233	AQ782619
17	35.4	5.0	667	236	AQ991044
18	35.4	5.0	819	146	BF269985
19	35.2	5.0	311	250	A2834870
20	35.2	5.0	590	250	A2834861
c	35	4.9	730	32	AV706549
22	34.8	4.9	295	161	BB565230
23	34.8	4.9	404	13	AA879865
24	34.8	4.9	459	12	AA794877
25	34.8	4.9	459	18	AI315494
26	34.8	4.9	479	8	AA497438
27	34.8	4.9	483	165	BE290613
28	34.8	4.9	499	1	AA065568
c	34.8	4.9	546	258	TA193A020
30	34.8	4.9	547	143	BF016664
31	34.8	4.9	552	166	BE303625
32	34.8	4.9	553	114	AW318849
33	34.8	4.9	585	166	BE371127
34	34.8	4.9	590	165	BE282182
35	34.8	4.9	590	168	BF719586
36	34.8	4.9	607	166	BE334491
37	34.8	4.9	692	166	BE377573
38	34.8	4.9	712	16	AI119048
39	34.8	4.9	712	235	AQ941109
40	34.8	4.9	860	145	BF167006
41	34.8	4.9	888	190	WR2115
42	34.8	4.9	926	145	BF140267
43	34.8	4.9	929	174	BF174120
44	34.8	4.9	936	169	BF788835
45	34.8	4.9	1301	192	AK014354

JOURNAL	Submitted (28-MAR-2001) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : segref@genoscope.cns.fr)				
REFERENCE	2 (bases 1 to 830)				
AUTHORS	Roth,C.W., Brey,P.T., Ke,Z., Collins,F.H. and Weissenbach,J.				
TITLE	Direct Submission				
JOURNAL	Submitted (28-MAR-2001) BBMI, Institut Pasteur, 25, rue du Dr. Roux, Paris 75015, France				
COMMENT	This clone is from an A. gambiae BAC library provided by F.H. Collins and sequenced by Genoscope in collaboration with the Laboratory of Biochem. and Biol. Molec. of Insects, Institut Pasteur.				
FEATURES	Location/Qualifiers				
source	1. 830				
BASE COUNT	207 a	231 c	227 g	156 t	9 others
ORIGIN					
Query Match	8.5%; Score 60.4; DB 219; Length 830;				
Best Local Similarity	50.7%; Pred. No. 8.5e-08;				
Matches	142; Conservative	1; Mismatches	137; Indels	0; Gaps	0;
QY	43 ggcgctccacctgttgcctgagcgatcgtcttgcgtccatgacgcaaacacaaag 102				
Db	318 GCCAGTTCGAGGTACTGCTCATCGCGTTCAGCGCGCGGATCCGTACCAGAGCAGA 377				
QY	103 taagcgatgcacccagcgcaataatcagcagcagcgcgcgcttggaacaaaga 162				
Db	378 AACGCCAGCGCGCCAGGTGACCAACGCCAGCAGCAGCAGCATCAGCAGCGCG 437				
QY	163 tcaacgccccaggtgcgcgatgacaaagagcgtcagaataataacacagagaaga 222				
Db	438 CTGCCGCCAAAATCCCGCAGATCAGCAGCAAAATCACTACCGCGCATAGCAGGGCA 497				
QY	223 accgcatgactcctgcgcttaattcctgttgaatcaccagtcattctcggtccg 282				
Db	498 ATCATGAGTGGTACTGCGCGGATCCCTGATTCATCAGACGCGATTTTCGGGCCG 557				
QY	283 atgacagtaaaagactggcccccaaaagcagacactgttaa 322				
Db	558 AGGGAGAGATATAAAATGCACCTAAGCAAGCCCTTGAM 597				
RESULT 2					
CNS015QU/c	916 bp DNA GSS 26-JUL-1999				
LOCUS	Drosophila melanogaster genome survey sequence SP6 end of BAC				
DEFINITION	BACN14D20 of DrosBAC library from Drosophila melanogaster (fruit fly); genomic survey sequence.				
ACCESSION	AL105792				
VERSION	AL105792.1 GI:5618347				
KEYWORDS	GSS.				
SOURCE	fruit fly.				
ORGANISM	Plasmod Drosophila melanogaster				
	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.				
REFERENCE	1 (bases 1 to 916)				
AUTHORS	Genoscope.				
TITLE	Direct Submission				
JOURNAL	Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : segref@genoscope.cns.fr)				
COMMENT	Determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) - http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billaud at CEPH (Centre				

[illegible]


```

Query Match          5.1%; Score 36; DB 228; Length 436;
Best Local Similarity 60.6%; Pred. No. 3.2;
Matches 57; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

Qy 614 tgggttgctgcaagcagaaaactaaacgcaactatctggcgcgcttgctgaaatccgct 673
      ||||| || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 187 TGGTGGTGCNACTAGCACAGAATAATGAGAAAGTTTCTGAAAGAGGTGGAGCCATCCTCT 128

Qy 674 taaccatgcgcatacaagcagagattcgctatccac 707
      | ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 127 TCCCATCCCATCTTCCCTCCATCNCATCCCC 94

RESULT 10
BF485823
LOCUS BF485823 650 bp mRNA EST 16-APR-2001
DEFINITION At19875.5:prime AT Drosophila melanogaster adult testes potB7
Drosophila melanogaster cDNA clone AT19875 5 similar to CGI3768:
F80013768 last updated:000321, mRNA sequence.
ACCESSION BF485823
VERSION BF485823.1 GI:11569124
KEYWORDS EST.
SOURCE Drosophila melanogaster
ORGANISM fruit fly.
REFERENCE 1
AUTHORS Stapleton,M., Brokstein,P., Hong,L., Agbayani,A., Baxter,E., Berman,
B., Carlson,J., Champe,M., Chavez,C., Chew,M., Dorsett,V., Farfan,
D., Frise,E., George,R., Gonzalez,M., Guarin,H., Harris,N., Li,P.,
Liao,G., Miranda,A., Mistra,S., Mungall,C.J., Nunoo,J., Pacleb,J.,
Park,S., Paragas,V., Phouanavong,S., Wan,K., Yu,C., Lewis,S.E.,
Celnaker,S. and Rubin,G.M.
TITLE Berkeley Drosophila Gene Collection Project
JOURNAL Unpublished (2000)
COMMENT Contact: Stapleton, M.
BDGP
Lawrence Berkeley National Lab
One Cyclotron Rd, Berkeley, CA 94720, USA
Fax: 510 486 6798
Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
hit genomic sequence AE003614
Plate: AT.198 row: G column: 3
High quality sequence stop: 633.
FEATURES
source
1..650
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="AT19875"
/clone.lib="AT Drosophila melanogaster adult testes potB7"
/sex="male"
/dev_stage="0-3 day old Ore-R males"
/lab_host="Plates AT.10-AT.120: DH5-alpha. Plates
AT.121-AT.319: DH5-alpha Tona"
/notes="Organ: ADULT testes; Vector: potB7; Site_1: ECORI;
Site_2: XhoI; The mRNA for the testis library was made
from testes and seminal vesicles hand dissected from 0-3
day old Ore-R males. RNA kindly provided by the lab of
Margaret Fuller. Sized fractionated cDNAs were directly
ligated into potB7. Plasmid cDNA library."
BASE COUNT 208 a 173 c 145 g 124 t
ORIGIN

Query Match          5.0%; Score 35.8; DB 149; Length 650;
Best Local Similarity 65.8%; Pred. No. 4.2;
Matches 52; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

Qy 421 ctggacacttgctctcaatcatgatgaaggcagcttcgaagcgctcttagccctt 480
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 188 CTGACAAGTGTCTTCATTAATGACGAAGCAGCAGCAGGAGGCCCTCAACATA 247

```

```

Qy 481 tccattcccccctcgcg 499
    ||| | ||| |||
Db 248 TCCACGCACACTTGGCAG 266

RESULT 11
AW351850/c 166 bp mRNA EST 01-FEB-2000
DEFINITION RC1-CT0199-180999-012-A09 CT0199 Homo sapiens cDNA, mRNA sequence.
ACCESSION AW351850
VERSION AW351850.1 GI:6849563
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 166)
AUTHORS HCGP http://www.ludwig.org.br/ORESTES.
TITLE The FAPESP/LICR Human Cancer Genome Project
JOURNAL Unpublished (1999)
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?l1=RC1&l2=RC1-CT0199-
180999-012-A09&t3=1999-09-18&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 51
High quality sequence stop: 166.
Location/Qualifiers
1..166
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="CT0199"
/dev_stage="Adult"
/note="Organ: colon; Vector: puc18; Site_1: SmaI; Site_2:
SmaI; A mini-library was made by cloning products derived
from ORESTES PCR (U.S. Letters Patent application No. 196
,716 - Ludwig Institute for Cancer Research) profiles
into the pUC 18 vector. Reverse transcription of tissue
mRNA and cDNA amplification were performed under low
stringency conditions."
BASE COUNT 42 a 41 c 52 g 31 t
ORIGIN

Query Match 5.0%; Score 35.6; DB 115; Length 166;
Best Local Similarity 58.5%; Pred. No. 3.3;
Matches 62; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

Qy 12 caccgtgtgtcttcttcaatgatctgtgcttcacacctgttctgtcatgctgc 71
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 137 CACATGGGTGGTGTCTCTACAAATGGTCACAGCCCTACACCTCGTTCTTGTTCGCTT 78

Qy 72 ttctcgtgcatacagcgaaaccataacaggtaagcgatgccacc 117
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 77 GGATCCCGCCCTGTGACATTCGCCGACGATGTGAGTTATGCCAGCC 32

RESULT 12
BF352878 222 bp mRNA EST 22-NOV-2000
DEFINITION IL3-HT0619-100700-208-H08 HT0619 Homo sapiens cDNA, mRNA sequence.
ACCESSION BF352878
VERSION BF352878.1 GI:11311952
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 222)
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?l1=IL3&l2=IL3-HT0619-
100700-208-H08&t3=2000-07-10&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 60
High quality sequence stop: 91.
Location/Qualifiers
1..222
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="HT0619"
/dev_stage="Adult"
/note="Organ: head_neck; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
BASE COUNT 45 a 67 c 55 g 55 t
ORIGIN

Query Match 5.0%; Score 35.6; DB 147; Length 222;
Best Local Similarity 58.5%; Pred. No. 3.5;
Matches 62; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

Qy 12 caccgtgtgtcttcttcaatgatctgtgcttcacacctgttctgtcatgctgc 71
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5 CACCATGGGTGGTGTCTCTACAAATGGTCACAGCCCTCACCACCTTCTTCTTTCACCTT 64

Qy 72 ttctcgtgcatacagcgaaaccataacaggtaagcgatgccacc 117
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 65 GGATCCCGCCCTGTGACATTCGCCGACCAAGTGAGTGATGCCACCC 110

RESULT 13
AQ851405 393 bp DNA GSS 13-APR-2001
LOCUS LMAJFV1_lm40b11.y1 Leishmania major FV1 random genomic library
DEFINITION Leishmania major genomic clone LMAJFV1_lm40b11 5', DNA sequence.
ACCESSION AQ851405
VERSION AQ851405.1 GI:6056053
KEYWORDS GSS.
SOURCE Leishmania major.
ORGANISM Leishmania major
Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
Leishmania.
REFERENCE 1 (bases 1 to 393)
AUTHORS Akopyants,N.S., Clifton,S.W., Martin,J., Pape,D., Wylie,T., Li,L.,

```

Kissinger, J., Roos, D.S., Marra, M., Hillier, L., Chinwalla, A., Blistain, A., Schmitt, A., Person, B., Theising, B., Ritter, E., Ronko, I., Bennett, J., Cole, R., Underwood, K., Cardenas, M., Gibbons, M., Harvey, N., McCann, R., Tsagaris, R., Williams, T., Jackson, Y., Bowers, Y., Swaller, T., Waterston, R., Wilson, R., and Beverley, S.M.

A survey of the Leishmania major Friedlin strain V1 genome by shotgun sequencing: a resource for DNA

Mol. Biochem. Parasitol. 113 (2), 337-340 (2001)

Other_GSSs: lm40b11.x1

Contact: Akopyants, NS / Beverley, SM

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@wustl.wustl.edu

Library construction: Natalia S. Akopyants, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center

If using this information please cite:

N.S. Akopyants and S.M. Beverley 'A survey of the Leishmania major Friedlin strain V1 genome by shotgun sequencing' and the Washington University Genome Sequencing Center For information on obtaining clone material please contact: Natalia S. Akopyants Ph.D. (natalia@borcim.wustl.edu) and/or Stephen M. Beverley Ph.D. (beverley@borcim.wustl.edu)

Seq primer: -40RP from Gibco

Class: shotgun

High quality sequence stop: 357.

Location/Qualifiers

source

1. .393

/organism="Leishmania major"

/strain="Friedlin strain V1"

/db_xref="taxon:5664"

/clone.lib="LMJFV1_lm40b11"

/lab_host="TOP10 (Invitrogen)"

/note="Vector: pZero-2 (Invitrogen); Site_1: EcoRV;

Genomic DNA was isolated from stationary phase cells. For this library, DNA was sheared to give a tight size distribution of 1-1.5kb fragments, blunt-ended with T4 DNA polymerase, dephosphorylated with Shrimp Alkaline Phosphatase and ligated into pZero-2 vector's EcoRV site."

BASE COUNT 105 a 97 c 116 g 75 t

ORIGIN

Query Match 5.0%; Score 35.6; DB 234; Length 393;

Best Local Similarity 53.6%; Pred. No. 4.2;

Matches 74; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

Qy 564 ggcacaaagcaacgagcggtgaagtccttctgtgcaagcagcgagaaatggtgtgtct 623

Db 185 GTCCGTAGCGCAGTGTGCGAGAACCTCACCCCTCAAGCAGCGCAGTGTGCTGCGGCT 244

Qy 624 gcaagcagaactaaagcgcaactatcttgagcgcttctgtaaatcccggttaaccatgc 683

Db 245 ACTAGCAGTAATCGCAGCGCGAAAGCGGAGGATCTGTAAGACCTGTGGTTAGCCACGCC 304

Qy 684 catcaacgacagattcgct 701

Db 305 ACTAAGCGGAATCTGAT 322

RESULT 14

AW067177/c

LOCUS AW067177 492 bp mRNA EST 12-OCT-1999

DEFINITION 683016G08.x1 683 - 14 day immature embryo from Hake lab (HS) Zea mays cDNA, mRNA sequence.

ACCESSION AW067177

VERSION AW067177.1 GI:6022249

KEYWORDS EST.

SOURCE Zea mays.

ORGANISM Zea mays

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoidae; Andropogoneae; Zea.

1 (bases 1 to 492)

AUTHORS Walbot, V.

TITLE Maize ESTs from various cDNA libraries sequenced at Stanford University

JOURNAL Unpublished (1999)

COMMENT Contact: Walbot V

Department of Biological Sciences

Stanford University

855 California Ave, Palo Alto, CA 94304, USA

Tel: 650 723 2227

Fax: 650 725 8221

Email: walbot@stanford.edu

Plate: 683016 row: G column: 08.

Location/Qualifiers

source

1. .492

/organism="Zea mays"

/cultivar="B73"

/db_xref="taxon:4577"

/clone.lib="683 - 14 day immature embryo from Hake lab (HS)"

/tissue_type="embryo"

/dev_stage="14 days after pollination"

/lab_host="DH10B"

/note="Organ: embryo; Vector: pBKCMV (Stratagene's Zap Express); Site_1: XhoI; Site_2: EcoRI; Directionally cloned, 14 day immature embryo library created with Stratagene's Zap Express cDNA protocol."

BASE COUNT 125 a 97 c 154 g 116 t

ORIGIN

Query Match 5.0%; Score 35.6; DB 111; Length 492;

Best Local Similarity 51.9%; Pred. No. 4.4;

Matches 80; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

Qy 231 gagtcttcgcgcttaattcctgttttaatacaccagtcattctgcgctcgagacag 290

Db 399 GAGTCCCTCGTGTGTTCGCCCTTTTTCGCCCCCGGCGCTGCTTTTCAAGTTCCTG 340

Qy 291 taaagactggcccaaaagcagacctgttaataagagatttcattccatgaccatcgtagc 350

Db 339 TAAAGGGGACAAACCCCAAGCCCGCAGAGATATACGTTTTCGCCGCCCGCGTGCCA 280

Qy 351 ctatggaagtacttaagtataaatgattgtgtctt 384

Db 279 ATATCGATAACTCTGCGCAATCGCCAGTTCCT 246

RESULT 15

AZ253065

LOCUS AZ253065 494 bp DNA GSS 15-JUN-2000

DEFINITION RPI-23-94A11.TV RPI-23 Mus musculus genomic clone RPI-23-94A11, DNA sequence.

ACCESSION AZ253065

VERSION AZ253065.1 GI:8566268

KEYWORDS GSS.

SOURCE house mouse.

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 494)

AUTHORS Zhao, S., Nierman, W., Feldblyum, T., Malek, J., Shatsman, S., Akinret, B., Levins, M., McGann, S., Tsegaye, G., Geer, K., Krol, M., de Jong, P., and Fraser, C.M.

TITLE Mouse BAC End Sequences from Library RPI-23

JOURNAL Unpublished (1999)

COMMENT Other_GSSs: RPI-23-94A11.TJ

Contact: Shaying Zhao

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: szhae@tigr.org
 Clones are derived from the mouse BAC library RPCI-23. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (<http://bacpac.med.buffalo.edu/orderingframe.htm>) or from Resea ch Genetics (info@resgen.com). BAC end page: http://www.tigr.org/tldb/bac_ends/mouse/bac_end_intro.html
 Plate: 94 row: A column: 11

Seq primer: T7
 Class: BAC ends.
 Location/Qualifiers
 1. .494
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone_lib="RPCI-23-94A11"
 /clone_lib="RPCI-23"
 /sex="Female"
 /lab_host="DH10B"
 /note="Organ: Kidney/Brain; Vector: pBACE3.6; Site_1: EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or brain genomic DNA was isolated and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBACE3.6 vector at the EcoRI sites. The ligation products were transformed into DH10B electrocompetent cells (BRL Life Technologies)."
 BASE COUNT 220 a 71 c 73 g 130 t
 ORIGIN

Query Match 5.0%; Score 35.6; DB 240; Length 494;
 Best Local Similarity 50.6%; Pred. NO. 4.4;
 Matches 86; Conservative 0; Mismatches 84; Indels 0; Gaps 0;
 QY 241 cgcttaattccttggttaataccaggtacattctgcggtccggtgacagtaaaagactg 300
 Db 318 CTCTAAACATTTTAAAAAAGTCTATTTCCTCCCAATGAAGATGATAAATCCACCT 377
 QY 301 gcccccaaaagacgtgtaagaagatttccatgataccatcgtgacctatggaagt 360
 Db 378 AATCCTATAACAAAGACCAACAATGATTACCATGCTTGACATAATGAAATAGTCAAAAT 437
 QY 361 actaagtaaatgattggttcttaacatggtttatagcttcattgaa 410
 Db 438 TATATACAAAATANTTATGCAATACCTAAATCAAAATCACTGGAGAA 487

Search completed: September 14, 2001, 06:59:12
 Job time: 12059 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 14, 2001, 05:26:31 ; Search time 171.99 Seconds
(without alignments)
2595.721 Million cell updates/sec

Title: us-09-105-117i-1-copy_1016_1726
Perfect score: 711
Sequence: 1 ggtcatcgggcacggttgg.....cagattcgctatccacatgg 711

Scoring table:
IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 730101 seqs, 313950809 residues

Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_0601.*
1: /cgnl_9/gcgdata/geneseq/geneseq/NA1980.DAT.*
2: /cgnl_9/gcgdata/geneseq/geneseq/NA1981.DAT.*
3: /cgnl_9/gcgdata/geneseq/geneseq/NA1982.DAT.*
4: /cgnl_9/gcgdata/geneseq/geneseq/NA1983.DAT.*
5: /cgnl_9/gcgdata/geneseq/geneseq/NA1984.DAT.*
6: /cgnl_9/gcgdata/geneseq/geneseq/NA1985.DAT.*
7: /cgnl_9/gcgdata/geneseq/geneseq/NA1986.DAT.*
8: /cgnl_9/gcgdata/geneseq/geneseq/NA1987.DAT.*
9: /cgnl_9/gcgdata/geneseq/geneseq/NA1988.DAT.*
10: /cgnl_9/gcgdata/geneseq/geneseq/NA1989.DAT.*
11: /cgnl_9/gcgdata/geneseq/geneseq/NA1990.DAT.*
12: /cgnl_9/gcgdata/geneseq/geneseq/NA1991.DAT.*
13: /cgnl_9/gcgdata/geneseq/geneseq/NA1992.DAT.*
14: /cgnl_9/gcgdata/geneseq/geneseq/NA1993.DAT.*
15: /cgnl_9/gcgdata/geneseq/geneseq/NA1994.DAT.*
16: /cgnl_9/gcgdata/geneseq/geneseq/NA1995.DAT.*
17: /cgnl_9/gcgdata/geneseq/geneseq/NA1996.DAT.*
18: /cgnl_9/gcgdata/geneseq/geneseq/NA1997.DAT.*
19: /cgnl_9/gcgdata/geneseq/geneseq/NA1998.DAT.*
20: /cgnl_9/gcgdata/geneseq/geneseq/NA1999.DAT.*
21: /cgnl_9/gcgdata/geneseq/geneseq/NA2000.DAT.*
22: /cgnl_9/gcgdata/geneseq/geneseq/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
C 1	711	100.0	2374	18 AAT96816	DNA encoding LysG,
C 2	435	61.2	822	22 AAF71779	Corynebacterium q1
C 3	406	57.1	993	22 AAF71777	Corynebacterium q1
4	63.2	8.9	15239	17 AAT33536	BCG deletion regio
C 5	60.4	8.5	636	21 AAT52891	Escherichia coli y
C 6	45.4	6.4	45613	22 AAF28535	Genomic fragment #
C 7	45.2	6.4	4915	20 AAX84317	Stealth virus nucl
8	37.8	5.3	936	22 AAF58252	Oligonucleotide D1
9	37.8	5.3	936	22 AAF58254	Oligonucleotide D1
10	37.8	5.3	936	22 AAF58257	Oligonucleotide D1
11	37.8	5.3	936	22 AAF58259	Oligonucleotide D2

12	37.8	5.3	936	22 AAF58262	Oligonucleotide D2
13	37.8	5.3	938	22 AAF58255	Oligonucleotide D1
C 14	37	5.2	2058	9 AAN80093	Sequence of BamHI/
C 15	37	5.2	2058	16 AAQ97857	P. aeruginosa tfda
16	37	5.2	102634	21 AAA81464	N. meningitidis pa
17	37	5.2	349980	21 AAF21609	Neisseria meningit
18	36.6	5.1	10732	21 AAA10594	Gene encoding a su
C 19	36	5.1	936	22 AAF58252	Oligonucleotide D1
C 20	36	5.1	936	22 AAF58254	Oligonucleotide D1
C 21	36	5.1	936	22 AAF58257	Oligonucleotide D1
C 22	36	5.1	936	22 AAF58259	Oligonucleotide D2
C 23	36	5.1	936	22 AAF58262	Oligonucleotide D2
C 24	36	5.1	938	22 AAF58255	Oligonucleotide D1
C 25	34.8	4.9	1431	8 AAT70264	Sequence encoding
26	34.6	4.9	245	21 AAC22623	Human secreted pro
27	34.6	4.9	1704	20 AAX99698	Human TGF-beta act
28	34.6	4.9	2785	20 AAX99696	Human TGF-beta act
29	34.6	4.9	2866	20 AAX99697	Human TGF-beta act
30	34.4	4.8	5819	20 AAX20533	Polynucleotide seq
31	33	4.6	1788	20 AAX56285	Human TAK1-6xHis e
32	33	4.6	1959	18 AAT85095	Human transforming
33	33	4.6	2656	20 AAX56279	Human TAK1 encodin
34	33	4.6	2656	21 AAA39105	Human TAK-1 nucleo
35	32.8	4.6	1876	21 AAA26882	Essential Staphylo
36	32.8	4.6	1876	22 AAF91564	Staphylococcus aur
37	32.6	4.6	346	21 AAA43240	Xenopus secreted e
38	32.6	4.6	1134	21 AAA24355	Neisseria meningit
39	32.6	4.6	1134	21 AA254489	Neisseria meningit
40	32.6	4.6	25509	21 AAA81467	N. meningitidis pa
C 41	32.6	4.6	349980	21 AAF21607	Neisseria meningit
C 42	32.4	4.6	3333	21 AAA93618	Human semaphorin p
C 43	32.4	4.6	3498	21 AAA93617	Human semaphorin p
C 44	32.4	4.6	3550	19 AAV44295	Human secreted pro
C 45	32.4	4.6	3862	21 AAD01233	Human semaphorin 6

ALIGNMENTS

RESULT 1
AAT96816/c
ID AAT96816 standard; DNA; 2374 BP.
XX AC AAT96816;
XX DT 12-MAR-1998 (first entry)
XX XX DNA encoding LysG, LysE and ORF3 from Corynebacterium glutamicum.
DE XX LysG; LysE; ORF3; lysine transport; regulatory protein; export protein;
KW KW Microbial production; amino acid; animal feed additive; ds.
XX XX Corynebacterium glutamicum.
OS
XX FH Key Location/Qualifiers
FT CDS complement (82..954)
FT FT /*tag= a
FT FT /label= LysG
FT FT 1016..1726
FT FT /*tag= b
FT FT /label= LysE
FT FT complement (1723..2373)
FT FT /*tag= c
FT FT /label= orf3
XX DE19548222-A1.
XX PN
XX XX
XX PD 26-JUN-1997.
XX XX
XX PF 22-DEC-1995; 95DE-1048222.
XX XX
XX PR 22-DEC-1995; 95DE-1048222.
XX XX

PA (KERJ) FORSCHUNGSZENTRUM JUELICH GMBH.

XX Eggeling L, Sahm H, Vrljic M;

XX WPI; 1997-333867/31.

DR P-PSDB; AAW377/14-16.

XX

PT Increasing microbial production of amino acids, especially lysine -
PT by improving export carrier activity or corresponding gene
PT expression, also new export and regulatory genes from
PT Corynebacterium

XX

PS Claim 23 and 26; Page -: 16pp; German.

XX

CC This DNA, isolated from Corynebacterium glutamicum, contains the LysG,
CC LysE and OmpF genes. LysG and LysE encode a lysine transport regulator
CC protein and an export protein, respectively. Microbial production of
CC amino acids (A) is improved by increasing the export-carrier activity
CC and/or the export gene expression in a microorganism that produces (A).
CC The method is specifically used to increase production of lysine,
CC used as an animal feed additive. Other (A) are variously useful as
CC pharmaceuticals, condiments and intermediates for fine chemicals.
CC This method increases the amount of (A) secreted into the culture medium.
CC Export of (A) has been found to depend on a single gene.
CC NB. This sequence has been created from the information given in table 2
CC of the specification.

XX

SQ Sequence 2374 BP; 526 A; 640 C; 648 G; 560 T; 0 other;

Query Match

Best Local Similarity 100.0%; Score 711; DB 18; Length 2374;

Matches 711; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gtgtcatcgaggcaggttggtctgttttcttcaatgatgtgtggtccaccctgttt 60
DB 1359 GTGTCAATCGGCGACGGTGGTGTGTTTCTTCAATGATGTGTGGCGCTTCCACCTGTGTTT 1300
QY 61 gtcatggcgtcttctcgtcgcacgacggaacacataacaggttaagcagtcacaccccg 120
DB 1299 GTCATGGCGTCTTTCGCTGCCATGACGGCAACCATACAGGTAAAGCGATGCCACCCAG 1240
QY 121 cgataatctcgagcagatcgcgcggttgacacacacacacacacacacacacacacacac 180
DB 1239 CGCATTAATTCGAGCAGATCGCGCGGCGATTTGACAAAAGATCAACGCCCAAGGTGCGG 1180
QY 181 gcatgaac 240
DB 1179 GCGATGAACAAAAGACGTCAGAAATTAACACACGAGAGAACCCGCAATGAGTCTTCG 1120
QY 241 cgcttaattctctgttcaatcaccagttacattctgcggtccgatgacacagtaaaagactg 300
DB 1119 CGCTTAATTCCTGTGTTTAAATCACCAGTACATCTCGCGTCCGATGGACAGTAAAGACTG 1060
QY 301 gcccccaagacagccttaataagatttcacatcaccatcgaccatcgacatggaagt 360
DB 1059 GCCCCCAAGACAGCCTTAATGAAGATTTCCATGATCACCATCGTACCTATGGAAGT 1000
QY 361 actaagtaaaatgattgtttcttaacatggttttaatatagcttaagccccattcaa 420
DB 999 ACTTAAGTAAATGATGTGTCTTAAATGTTTAAATAGCTTTCATGAACCCCAATCAA 940
QY 421 ctggacatttgcctcaatcatgatgagcgagcttcgaagcgccctccttagccctt 480
DB 939 CTGGACATTTGCTCTCATCATGATGATGAAGCAGCTTCGAAGGCGCCTCTTAGCCCTT 880
QY 481 tcatctccccctcggtgagtcagcgttaagctctcagacatcacgtgggtcga 540
DB 879 TCATTTCCCTCCGCGGTGAGTCAGCGGTAAAGCTCTCAGCATCACGTGGGTGCA 820
QY 541 gtgttggtatcgac 600
DB 819 GTGTTGTTATCGCGCACCCCAACCGGCAAGCAACCGGAGGAGTCTTGTGCAA 760

QY 601 gcagcgcggaataatggtgtgtgcaagcagaaactaaagcgaactatctgagcgcctt 660
DB 759 GCAGCGCGGAAATGCTGTGTCGAAGCAGAACTAAAGCGCACTATCTGGACGCTT 700
QY 661 gctgaatacccttaaccatcgccatcgccatcgccatcgccatcgccatcgccatcgccatcg 711
DB 699 GCTGAATCCCGTTAACCATCGCCATCAACGAGATTCGCTATCCACATGG 649

RESULT 2

AAF71779/c

ID AAF71779 standard; DNA; 822 BP.

XX

AC AAF71779;

XX

DT 30-APR-2001 (first entry)

XX

DE Corynebacterium glutamicum MP protein nucleotide sequence SEQ ID NO:53.

XX

CC Corynebacterium glutamicum; metabolic pathway protein; MP protein;
KW fine chemical production; microorganism; organic acid; nucleoside;
KW nonproteinogenic amino acid; purine base; pyrimidine base; nucleotide;
KW lipid; saturated fatty acid; unsaturated fatty acid; diol; vitamin;
KW carbohydrate; aromatic compound; cofactor; polyketide; enzyme; ds.

XX

OS Corynebacterium glutamicum.

XX

PN WO200100843-A2.

XX

PD 04-JAN-2001.

XX

PF 23-JUN-2000; 2000WO-IB00923.

XX

PR 25-JUN-1999; 99US-0141031.

PR

PR 01-JUL-1999; 99DE-1030476.

PR

PR 02-JUL-1999; 99US-0142101.

PR

PR 08-JUL-1999; 99DE-1031415.

PR

PR 08-JUL-1999; 99DE-1031418.

PR

PR 08-JUL-1999; 99DE-1031419.

PR

PR 08-JUL-1999; 99DE-1031420.

PR

PR 08-JUL-1999; 99DE-1031424.

PR

PR 08-JUL-1999; 99DE-1031428.

PR

PR 08-JUL-1999; 99DE-1031434.

PR

PR 08-JUL-1999; 99DE-1031435.

PR

PR 08-JUL-1999; 99DE-1031443.

PR

PR 08-JUL-1999; 99DE-1031453.

PR

PR 08-JUL-1999; 99DE-1031457.

PR

PR 08-JUL-1999; 99DE-1031465.

PR

PR 08-JUL-1999; 99DE-1031478.

PR

PR 08-JUL-1999; 99DE-1031510.

PR

PR 08-JUL-1999; 99DE-1031541.

PR

PR 08-JUL-1999; 99DE-1031573.

PR

PR 08-JUL-1999; 99DE-1031592.

PR

PR 08-JUL-1999; 99DE-1031632.

PR

PR 08-JUL-1999; 99DE-1031634.

PR

PR 08-JUL-1999; 99DE-1031636.

PR

PR 09-JUL-1999; 99DE-1032125.

PR

PR 09-JUL-1999; 99DE-1032126.

PR

PR 09-JUL-1999; 99DE-1032130.

PR

PR 09-JUL-1999; 99DE-1032186.

PR

PR 09-JUL-1999; 99DE-1032206.

PR

PR 09-JUL-1999; 99DE-1032227.

PR

PR 09-JUL-1999; 99DE-1032228.

PR

PR 09-JUL-1999; 99DE-1032229.

PR

PR 09-JUL-1999; 99DE-1032230.

PR

PR 14-JUL-1999; 99DE-1032922.

PR

PR 14-JUL-1999; 99DE-1032926.

PR

PR 14-JUL-1999; 99DE-1032928.

PR

PR 14-JUL-1999; 99DE-1033004.

PR

PR 14-JUL-1999; 99DE-1033005.

PR

PR 14-JUL-1999; 99DE-1033006.

PR

PR 12-AUG-1999; 99US-0148613.

PR 27-AUG-1999; 99DE-1040764.
 PR 27-AUG-1999; 99DE-1040765.
 PR 27-AUG-1999; 99DE-1040766.
 PR 27-AUG-1999; 99DE-1040832.
 PR 31-AUG-1999; 99DE-1041378.
 PR 31-AUG-1999; 99DE-1041379.
 PR 31-AUG-1999; 99DE-1041380.
 PR 31-AUG-1999; 99DE-1041394.
 PR 31-AUG-1999; 99DE-1041396.
 PR 03-SEP-1999; 99DE-1042076.
 PR 03-SEP-1999; 99DE-1042077.
 PR 03-SEP-1999; 99DE-1042079.
 PR 03-SEP-1999; 99DE-1042086.
 PR 03-SEP-1999; 99DE-1042087.
 PR 03-SEP-1999; 99DE-1042088.
 PR 03-SEP-1999; 99DE-1042095.
 PR 03-SEP-1999; 99DE-1042124.
 PR 03-SEP-1999; 99DE-1042129.
 PR 09-MAR-2000; 2000US-0187970.
 XX (BADI) BASF AG.
 PA Pompejus M, Kroeger B, Schroeder H, Zelder O, Haberhauer G;
 PI WPI; 2001-137957/14.
 DR P-PSDB; AAB79660.
 DR
 XX Nucleic acids from Corynebacterium glutamicum encoding metabolic
 PT pathway proteins, useful for producing fine chemicals in
 PT microorganisms, including organic acids, nonproteinogenic amino acids,
 PT and purine and pyrimidine bases -
 XX
 PS Claim 3; Page 233-234; 1737pp; English.
 XX
 CC AAF71753 to AAF72330 encode the Corynebacterium glutamicum metabolic
 CC pathway (MP) proteins given in AAB79634 to AAB80211. The C. glutamicum
 CC MP nucleic acids are useful for the production of fine chemicals
 CC in microorganisms, including organic acids, nonproteinogenic amino
 CC acids, purine and pyrimidine bases, nucleosides, nucleotides, lipids,
 CC saturated and unsaturated fatty acids, diols, carbohydrates, aromatic
 CC compounds, vitamins, cofactors, polyketides and enzymes.
 XX
 SQ Sequence 822 BP; 167 A; 192 C; 246 G; 217 T; 0 other;

Query Match 61.2%; Score 435; DB 22; Length 822;
 Best Local Similarity 100.0%; Pred. No. 8e-133;
 Matches 435; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gtgtcatcgagcaggttggtctgtttcttcaatgatctgtgagcttcacattgttt 60
 Db 435 GTGTCATCGGCGACGGTTGGTTCGTCTTCTTCAATGATCTGTGGCGCTTCCACCTTGT 376
 Qy 61 gtcatgcttcttcgtccatcacgcaacacataacagtaagcagatccaccaccag 120
 Db 375 GTCATGCGCTTCTTCGCTGCCATACGGCAACCATACAGTAAGCGATGCCACCCAG 316
 Qy 121 cgcataatcgacagcagtcgctgagcattgacaaaagatcaacgcccaagtgccg 180
 Db 315 CGCATATATCGACAGCATCGCGCGCATTCGACAAAAGATCAACGCCCAAGTGCCG 256
 Qy 181 gcgatacaaaaaagcgtcagaattaaacacacacgagaagaacgcgaatgagtcctcg 240
 Db 255 CGCATGAACAAAAGCGTCAGAAATTAACACACGACGAGAACCGCAATGAGTCCTCG 196
 Qy 241 cgttaattctctttaaataacacagtaacattctgcggtccgagtgacagtaaaagactg 300
 Db 195 CGCTTAATTCCTTTTAAATCACAGTACATTCCTGCGGTCCGATGGACAGTAAAGACTG 136
 Qy 301 gcccccaaacagcacctgtaataagatttccatgatcacccatcgtgacctatggaagt 360
 Db 135 GCCCCAAACGACACCTGTAAAGAGATTTCATGATCACCATCGTGACCTATGGAAGT 76

Qy 361 acttaagtaaatgatgtggttcttcaacatggttcaataatagcttcacgaaccccatcca 420
 Db 75 ACTTAAGTAAATGATGTGGTTCTTAACATGTTTAAATATAGCTTCATGACCCCATTCAA 16
 Qy 421 ctggacacatttgc 435
 Db 15 CTGGACACTTTGCTC 1
 RESULT 3
 AAF71777
 ID AAF71777 standard; DNA; 993 BP.
 XX
 AC AAF71777;
 XX
 DT 30-APR-2001 (first entry)
 XX
 DE Corynebacterium glutamicum MP protein nucleotide sequence SEQ ID NO:49.
 XX
 KW Corynebacterium glutamicum; metabolic pathway protein; MP protein;
 KW fine chemical production; microorganism; organic acid; nucleoside;
 KW nonproteinogenic amino acid; purine base; pyrimidine base; nucleotide;
 KW lipid; saturated fatty acid; unsaturated fatty acid; diol; vitamin;
 KW carbohydrate; aromatic compound; cofactor; polyketide; enzyme; ds.
 XX
 OS Corynebacterium glutamicum.
 XX
 PN WO200100843-A2.
 XX
 PD 04-JAN-2001.
 XX
 PF 23-JUN-2000; 2000WO-IB00923.
 XX
 PR 25-JUN-1999; 99US-0141031.
 PR 01-JUL-1999; 99DE-1030476.
 PR 02-JUL-1999; 99US-0142101.
 PR 08-JUL-1999; 99DE-1031415.
 PR 08-JUL-1999; 99DE-1031418.
 PR 08-JUL-1999; 99DE-1031419.
 PR 08-JUL-1999; 99DE-1031420.
 PR 08-JUL-1999; 99DE-1031424.
 PR 08-JUL-1999; 99DE-1031428.
 PR 08-JUL-1999; 99DE-1031434.
 PR 08-JUL-1999; 99DE-1031435.
 PR 08-JUL-1999; 99DE-1031443.
 PR 08-JUL-1999; 99DE-1031453.
 PR 08-JUL-1999; 99DE-1031457.
 PR 08-JUL-1999; 99DE-1031465.
 PR 08-JUL-1999; 99DE-1031478.
 PR 08-JUL-1999; 99DE-1031510.
 PR 08-JUL-1999; 99DE-1031541.
 PR 08-JUL-1999; 99DE-1031573.
 PR 08-JUL-1999; 99DE-1031592.
 PR 08-JUL-1999; 99DE-1031632.
 PR 08-JUL-1999; 99DE-1031634.
 PR 08-JUL-1999; 99DE-1031636.
 PR 09-JUL-1999; 99DE-1032125.
 PR 09-JUL-1999; 99DE-1032126.
 PR 09-JUL-1999; 99DE-1032130.
 PR 09-JUL-1999; 99DE-1032186.
 PR 09-JUL-1999; 99DE-1032206.
 PR 09-JUL-1999; 99DE-1032227.
 PR 09-JUL-1999; 99DE-1032228.
 PR 09-JUL-1999; 99DE-1032229.
 PR 09-JUL-1999; 99DE-1032230.
 PR 14-JUL-1999; 99DE-1032922.
 PR 14-JUL-1999; 99DE-1032926.
 PR 14-JUL-1999; 99DE-1033004.
 PR 14-JUL-1999; 99DE-1033005.
 PR 14-JUL-1999; 99DE-1033006.
 PR 12-AUG-1999; 99US-0148613.
 PR 27-AUG-1999; 99DE-1040764.

PR 27-AUG-1999; 99DE-1040765.
 PR 27-AUG-1999; 99DE-1040766.
 PR 27-AUG-1999; 99DE-1040832.
 PR 31-AUG-1999; 99DE-1041378.
 PR 31-AUG-1999; 99DE-1041379.
 PR 31-AUG-1999; 99DE-1041380.
 PR 31-AUG-1999; 99DE-1041394.
 PR 31-AUG-1999; 99DE-1041396.
 PR 03-SEP-1999; 99DE-1042076.
 PR 03-SEP-1999; 99DE-1042077.
 PR 03-SEP-1999; 99DE-1042079.
 PR 03-SEP-1999; 99DE-1042086.
 PR 03-SEP-1999; 99DE-1042087.
 PR 03-SEP-1999; 99DE-1042088.
 PR 03-SEP-1999; 99DE-1042095.
 PR 03-SEP-1999; 99DE-1042124.
 PR 03-SEP-1999; 99DE-1042129.
 PR 09-MAR-2000; 2000US-0187970.
 XX
 PA (BADI) BASF AG.
 XX
 PI Pompejus M, Kroeger B, Schroeder H, Zelder O, Haberhauer G;
 XX
 DR WPI; 2001-137957/14.
 DR P-PSDB; AAB79658.
 XX
 XX Nucleic acids from Corynebacterium glutamicum encoding metabolic
 PT pathway proteins, useful for producing fine chemicals in
 PT microorganisms, including organic acids, nonproteinogenic amino acids,
 PT and purine and pyrimidine bases -
 XX
 PS Claim 3; Page 226-228; 1737pp; English.
 XX
 CC AAF71753 to AAF72330 encode the Corynebacterium glutamicum metabolic
 CC pathway (MP) proteins given in AAB79634 to AAB80211. The C. glutamicum
 CC MP nucleic acids are useful for the production of fine chemicals
 CC in microorganisms, including organic acids, nonproteinogenic amino
 CC acids, purine and pyrimidine bases, nucleosides, nucleotides, lipids,
 CC saturated and unsaturated fatty acids, diols, carbohydrates, aromatic
 CC compounds, vitamins, cofactors, polyketides and enzymes.
 XX
 SQ Sequence 993 BP; 233 A; 256 C; 268 G; 236 T; 0 other;

Query Match 57.1%; Score 406; DB 22; Length 993;
 Best Local Similarity 100.0%; Pred. No. 3.1e-123;
 Matches 406; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 306 caaaagcagactgtaatagaattccatgatccatcgacatcgtaagtgactta 365
 Db 1 caaaagcagactgtaatagaattccatgatccatcgacatcgtaagtgactta 60
 QY 366 agtaaatgatgtgtcttaacatggtttaataatagatcgaacccatcaactga 425
 Db 61 agtaaatgatgtgtcttaacatggtttaataatagatcgaacccatcaactga 120
 QY 426 cacttggctctcaatcattgatgaagcagcttcgaagcgctcttgccttccat 485
 Db 121 cacttggctctcaatcattgatgaagcagcttcgaagcgctcttgccttccat 180
 QY 486 ttccctcgcgggtgagtcagcggttaagctctcgagcatcagtggtgagtggt 545
 Db 181 ttccctcgcgggtgagtcagcggttaagctctcgagcatcagtggtgagtggt 240
 QY 546 ggtatcgacacccacccgcaaacgcaacccggaagcggtgagtggtgagtcagcagc 605
 Db 241 ggtatcgacacccacccgcaaacgcaacccggaagcggtgagtggtgagtcagcagc 300
 QY 606 gcggaataatggtgtgtctgcaagcagaaataaagcgcaactatctggcgcttctga 665
 Db 301 gcggaataatggtgtgtctgcaagcagaaataaagcgcaactatctggcgcttctga 360
 QY 566 aatcccggttaaccatcgccatcaacgcagattcgctatccaatgg 711

Db 361 aatcccggttaaccatcgccatcaacgcagattcgctatccaatgg 406
 RESULT 4
 AAT33536
 ID AAT33536 standard; DNA; 15239 BP.
 XX
 AC AAT33536;
 XX
 DT 15-FEB-1998 (first entry)
 XX
 DE BCG deletion region 2 and flanking sequences.
 KW BCG delta 2; virulence; avirulence; attenuation; gene deletion;
 KW mycobacteria; vaccine; infection; marker; ss.
 XX
 OS Mycobacterium bovis strain BCG.
 XX
 FH Key Location/Qualifiers
 FT misc_feature 3382..14071
 FT /*tag= a
 FT /note= "BCG delta 1 deletion region"
 XX
 PN W09625519-A1.
 XX
 PD 22-AUG-1996.
 XX
 PF 15-FEB-1996; 96WO-US01938.
 XX
 PR 17-FEB-1995; 950S-0390878.
 XX
 PA (PATH-) PATHOGENESIS CORP.
 XX
 PI Mahairas GG, Stover CK;
 XX
 DR WPI; 1996-393419/39.
 XX
 PT Detecting markers for avirulence in Mycobacterium - used in
 PT production of vaccines against bacterial infection, and to detect
 PT bacterial infection
 XX
 PS Example 1; Fig 2; 66pp; English.
 XX
 CC This DNA sequence comprises Mycobacterium bovis BCG deletion
 CC sequence BCGdelta2. A specific genetic deletion of this region
 CC results in an avirulence phenotype of the mycobacterium. 2 Other
 CC deletion regions (see AAT33535 and AAT33537) have also been detected.
 CC Identification involved screening a BCG cosmid library with a
 CC radiolabeled probe obtained following DNA subtraction between
 CC virulent Mycobacterium tuberculosis H37Rv and avirulent BCG.
 CC The deletions provide useful markers for the identification of an
 CC avirulent, or a virulent, mycobacterial phenotype. Determination
 CC of avirulence requires the detection of the presence or absence of
 CC the deletion; the deletions are detected either by detecting the
 CC presence or absence of deletion junctions (see AAT33538-46), or by
 CC detecting the presence or absence of the sequences contained within
 CC the deletion. Deletion polypeptides are used as components of
 CC immunological assays and in vaccines.
 XX
 SQ Sequence 15239 BP; 2878 A; 4791 C; 4729 G; 2841 T; 0 other;

Query Match 8.9%; Score 63.2; DB 17; Length 15239;
 Best Local Similarity 55.3%; Pred. No. 7.4e-10;
 Matches 166; Conservative 0; Mismatches 128; Indels 6; Gaps 2;

QY 418 caactggacacttgcctcctcaatcattgatgaagcagcttcgaagcgctcttagcc 477
 Db 5253 cagctggcgccgcttgcctcctcctcaatcattgatgaagcagcttcgaagcgctcttagcc 5312
 QY 478 ctttccatttccccctggcgctgagtcagcggttaagctctgagcatcacgtgggt 537

Query Match 8.5%; Score 60.4; DB 21; Length 636;
Best Local Similarity 51.1%; Pred. No. 1e-09;
Matches 142; Conservative 0; Mismatches 136; Indels 0; Gaps 0;

Query Match .6.4%; Score 45.4; DB 22; Length 45613;
Best Local Similarity 57.3%; Pred. No. 0.00098;

Db	1461	CACCTCTTGGCTGCTGGTGGTACGACGAGGACTCTCAGCAGGTGCAGCGCATGAACCTCCGCAT	1402
Qy	486	ttccctccctcgagcgtagtcagcgcggttaaagctctcagacatcacgctgggctcgagtggtt	545
Db	1401	TACTCCCTCCGCTGTGAGCCAGCGGCTCAAGGCTTTGGAGACTCTTCTTGGTCAGGTCGT	1342
Qy	546	gglatcgccgacacccacccggccaaagcaaccgaacgcggggtgaagtcccttgtgcaagcagc	605
Db	1341	ATTGGTGGCTCAAAAGCCAGTCCGCTGACGAGCAGCGGGGGCTCCATCGTTCGGTTTGC	1282
Qy	606	gcggaaaaatggttgtgtcgtcaagcagaaac	635
Db	1281	TCGCTCAGACCGACATGCTCCAGACGGGAGAC	1252

RESULT	8
AAF58252	
ID	AAF58252 standard; DNA; 936 BP.
XX	
XX	AAF58252;
XX	
DT	24-APR-2001 (first entry)
XX	
DE	Oligonucleotide D1835.
XX	
KW	Electron-transfer group; ETM; mismatch; genotyping;
KW	gene expression; ss.
XX	
OS	Synthetic.

[illegible]

[illegible]

	Matches	18; Conservative	163; Mismatches	130; Indels	0; Gaps
Qy	184	atgaacaaaagcgtcagaaa	taaacacacgagagaacgc	aatgagtccttcgcgc	243
	:	:	:	:	:
	:	:	:	:	:
	:	:	:	:	:
Db	512	wwwwwwwwwwwwwwwwww	wwwwwwwwwwwwwwwwww	wwwwwwwwwwwwwwwwww	571
Qy	244	ttaatctctgtttaatca	accagtagcattctgcggt	tcgatgacacagtaaaagac	gtgcgc 303
	:	:	:	:	:
	:	:	:	:	:
	:	:	:	:	:
Db	572	wwwwwwwwwwwwwwwwww	wwwwwwwwwwwwwwwwww	wwwwwwwwwwwwwwwwww	631
Qy	304	cccaaaaagcagacgtg	taatgaagattccatgat	caccatcgtgacctatgga	agtact 363
	:	:	:	:	:
	:	:	:	:	:
	:	:	:	:	:
Db	632	wwwwwwwwwwwwwwwwww	wwwwwwwwwwwwwwwwww	wwwwwwwwwwwwwwwwww	691
Qy	364	taagtaaatgattggtct	ttaacatggtttaatat	tagcttcacgaaccccat	tcaactg 423
	:	:	:	:	:
	:	:	:	:	:
	:	:	:	:	:
Db	592	wwwwwwwwwwwwwwwwww	wwwwwwwwwwwwwwwwww	wwwwwwwwwwwwwwwwww	751
Qy	424	gacacttgctccaatc	attgatgaaggcagctgc	gaaggcgccctcttagc	cttttcc 483
	:	:	:	:	:
	:	:	:	:	:
	:	:	:	:	:
Db	752	wwwwwwwwwwwwwwwwww	wwwwwwwwwwwwwwwwww	wwwwwwwwwwwwwwwwww	811
Qy	484	atttccccctc	494		
Db	812	cccccccccc	822		

RESULT 14
AAN80093/c
ID AAN80093 standard; DNA: 2058 BP.

DE Oligonucleotide D1876.

sequence of BamHI/Sali fragment of the *tfdA* gene which encodes 2,4-dichlorophenoxyacetic acid (2,4-D) mono-oxygenase.

KW Herbicide resistance; halogenated aromatic cpd; enzyme;
KW mono-oxygenase; SS.

OS *Alcaligenes eutrophus* JMP134.

XX	Key	Location/Qual
FH		

FM	KEY	LOCATION/Qualifiers
FT	CDS	748..1608

```

Fm1  /tag= a
Fm2  /note= "DNA so claimed"

```

XX
DN
DE2620000-7

XX 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015 2016 2017 2018 2019 2020 2021 2022 2023 2024 2025 2026 2027 2028 2029 2030 2031 2032 2033 2034 2035 2036 2037 2038 2039 2040 2041 2042 2043 2044 2045 2046 2047 2048 2049 2050 2051 2052 2053 2054 2055 2056 2057 2058 2059 2060 2061 2062 2063 2064 2065 2066 2067 2068 2069 2070 2071 2072 2073 2074 2075 2076 2077 2078 2079 2080 2081 2082 2083 2084 2085 2086 2087 2088 2089 2090 2091 2092 2093 2094 2095 2096 2097 2098 2099 2100 2101 2102 2103 2104 2105 2106 2107 2108 2109 2110 2111 2112 2113 2114 2115 2116 2117 2118 2119 2120 2121 2122 2123 2124 2125 2126 2127 2128 2129 2130 2131 2132 2133 2134 2135 2136 2137 2138 2139 2140 2141 2142 2143 2144 2145 2146 2147 2148 2149 2150 2151 2152 2153 2154 2155 2156 2157 2158 2159 2160 2161 2162 2163 2164 2165 2166 2167 2168 2169 2170 2171 2172 2173 2174 2175 2176 2177 2178 2179 2180 2181 2182 2183 2184 2185 2186 2187 2188 2189 2190 2191 2192 2193 2194 2195 2196 2197 2198 2199 2200 2201 2202 2203 2204 2205 2206 2207 2208 2209 2210 2211 2212 2213 2214 2215 2216 2217 2218 2219 2220 2221 2222 2223 2224 2225 2226 2227 2228 2229 2230 2231 2232 2233 2234 2235 2236 2237 2238 2239 2240 2241 2242 2243 2244 2245 2246 2247 2248 2249 2250 2251 2252 2253 2254 2255 2256 2257 2258 2259 2260 2261 2262 2263 2264 2265 2266 2267 2268 2269 2270 2271 2272 2273 2274 2275 2276 2277 2278 2279 2280 2281 2282 2283 2284 2285 2286 2287 2288 2289 2290 2291 2292 2293 2294 2295 2296 2297 2298 2299 2300 2301 2302 2303 2304 2305 2306 2307 2308 2309 2310 2311 2312 2313 2314 2315 2316 2317 2318 2319 2320 2321 2322 2323 2324 2325 2326 2327 2328 2329 2330 2331 2332 2333 2334 2335 2336 2337 2338 2339 2340 2341 2342 2343 2344 2345 2346 2347 2348 2349 2350 2351 2352 2353 2354 2355 2356 2357 2358 2359 2360 2361 2362 2363 2364 2365 2366 2367 2368 2369 2370 2371 2372 2373 2374 2375 2376 2377 2378 2379 2380 2381 2382 2383 2384 2385 2386 2387 2388 2389 2390 2391 2392 2393 2394 2395 2396 2397 2398 2399 2400 2401 2402 2403 2404 2405 2406 2407 2408 2409 2410 2411 2412 2413 2414 2415 2416 2417 2418 2419 2420 2421 2422 2423 2424 2425 2426 2427 2428 2429 2430 2431 2432 2433 2434 2435 2436 2437 2438 2439 2440 2441 2442 2443 2444 2445 2446 2447 2448 2449 2450 2451 2452 2453 2454 2455 2456 2457 2458 2459 2460 2461 2462 2463 2464 2465 2466 2467 2468 2469 2470 2471 2472 2473 2474 2475 2476 2477 2478 2479 2480 2481 2482 2483 2484 2485 2486 2487 2488 2489 2490 2491 2492 2493 2494 2495 2496 2497 2498 2499 2500 2501 2502 2503 2504 2505 2506 2507 2508 2509 2510 2511 2512 2513 2514 2515 2516 2517 2518 2519 2520 2521 2522 2523 2524 2525 2526 2527 2528 2529 2530 2531 2532 2533 2534 2535 2536 2537 2538 2539 2540 2541 2542 2543 2544 2545 2546 2547 2548 2549 2550 2551 2552 2553 2554 2555 2556 2557 2558 2559 2560 2561 2562 2563 2564 2565 2566 2567 2568 2569 2570 2571 2572 2573 2574 2575 2576 2577 2578 2579 2580 2581 2582 2583 2584 2585 2586 2587 2588 2589 2590 2591 2592 2593 2594 2595 2596 2597 2598 2599 2600 2601 2602 2603 2604 2605 2606 2607 2608 2609 2610 2611 2612 2613 2614 2615 2616 2617 2618 2619 2620 2621 2622 2623 2624 2625 2626 2627 2628 2629 2630 2631 2632 2633 2634 2635 2636 2637 2638 2639 2640 2641 2642 2643 2644 2645 2646 2647 2648 2649 2650 2651 2652 2653 2654 2655 2656 2657 2658 2659 2660 2661 2662 2663 2664 2665 2666 2667 2668 2669 2670 2671 2672 2673 2674 2675 2676 2677 2678 2679 2680 2681 2682 2683 2684 2685 2686 2687 2688 2689 2690 2691 2692 2693 2694 2695 2696 2697 2698 2699 2700 2701 2702 2703 2704 2705 2706 2707 2708 2709 2710 2711 2712 2713 2714 2715 2716 2717 2718 2719 2720 2721 2722 2723 2724 2725 2726 2727 2728 2729 2730 2731 2732 2733 2734 2735 2736 2737 2738 2739 2740 2741 2742 2743 2744 2745 2746 2747 2748 2749 2750 2751 2752 2753 2754 2755 2756 2757 2758 2759 2760 2761 2762 2763 2764 2765 2766 2767 2768 2769 2770 2771 2772 2773 2774 2775 2776 2777 2778 2779 2780 2781 2782 2783 2784 2785 2786 2787 2788 2789 2790 2791 2792 2793 2794 2795 2796 2797 2798 2799 2800 2801 2802 2803 2804 2805 2806 2807 2808 2809 2810 2811 2812 2813 2814 2815 2816 2817 2818

XX XX

XX

PA (SCHD) SCHERING AG.
XX
XX
29 AUG 1980, 00DE 3043030.

[illegible]

FI SLEDGEI W, ILMUMIS AN, ZEIK W, XX
XX
DR WPI; 1988-071716/11.
DR P-PSDB; AAP80079.
XX
PT plasmids and bacteria contg. gene tfdA for 2,4-D-mono:oxygenase -
PT isolated using new Alcaligenes eutrophus transposon mutants

Disclosure: Fig 10: 32pp: German.

CC The *tfdA* gene product is a 2,4-D decomposing protein. A.eutrophus
CC JMP 134 contains an 80 kb plasmid, pJP4, which carries the genes
CC coding for 2,4-D decomposition. Plasmids and bacteria contg. the
XX

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 13, 2001, 17:33:09 ; Search time 34.57 Seconds
(without alignments)
378.790 Million cell updates/sec

Title: US-09-105-117i-2
Perfect score: 1091
Sequence: 1 DPLEETMTALRDIVASGRA.....SHDGINIWAKATDSKTREN 216

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 412676 seqs, 60623988 residues

Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_0601.*
1: /SIDS8/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SIDS8/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SIDS8/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SIDS8/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SIDS8/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SIDS8/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SIDS8/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SIDS8/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SIDS8/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SIDS8/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SIDS8/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SIDS8/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SIDS8/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SIDS8/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SIDS8/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SIDS8/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SIDS8/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SIDS8/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SIDS8/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SIDS8/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SIDS8/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SIDS8/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1091	100.0	216	18 AAW37716	C. glutamicum orf3
2	792	72.6	290	22 AAB76844	Corynebacterium gl
3	225	20.6	348	20 AAV52837	Escherichia coli p
4	205	18.8	329	18 AAW23388	Shaker-like potass
5	202	18.5	329	18 AAW23389	Shaker-like potass
6	195.5	17.9	266	21 AAG35920	Zea mays protein f
7	195.5	17.9	328	21 AAG35919	Zea mays protein f
8	195	17.9	216	21 AAG09266	Arabidopsis thalia
9	195	17.9	216	21 AAG39051	Arabidopsis thalia
10	195	17.9	266	21 AAG39050	Arabidopsis thalia
11	195	17.9	328	21 AAG39049	Arabidopsis thalia

12	184	16.9	312	22 AAB79618	Corynebacterium gl
13	184	16.9	312	22 AAB80067	Corynebacterium gl
14	180.5	16.5	315	22 AAB76845	Zea mays protein f
15	180.5	16.5	316	21 AAG35251	Zea mays protein f
16	180.5	16.5	371	21 AAG35250	Zea mays protein f
17	180.5	16.5	388	21 AAG35249	Sugar biosynthesis
18	175	16.0	333	18 AAW19736	S. erythraea drpp-4
19	175	16.0	333	20 AAW99386	Arabidopsis thalia
20	166.5	15.3	316	21 AAG39500	Arabidopsis thalia
21	166.5	15.3	324	21 AAG39499	Arabidopsis thalia
22	166.5	15.3	365	21 AAG39498	Arabidopsis thalia
23	164	15.0	192	21 AAG09267	Arabidopsis thalia
24	163.5	15.0	357	21 AAG41676	Arabidopsis thalia
25	163.5	15.0	375	21 AAG52050	Arabidopsis thalia
26	163.5	15.0	377	21 AAG52049	Arabidopsis thalia
27	163.5	15.0	386	21 AAG52048	Arabidopsis thalia
28	163.5	15.0	389	21 AAG52008	Arabidopsis thalia
29	163.5	15.0	412	21 AAG41675	Arabidopsis thalia
30	163.5	15.0	419	21 AAG41674	Arabidopsis thalia
31	162.5	14.9	331	20 AAY49837	Human aflatoxin B1
32	162.5	14.9	331	20 AAY49920	Human aflatoxin B1
33	162.5	14.9	331	21 AAB09902	Arabidopsis thalia
34	162	14.8	340	21 AAG38206	Arabidopsis thalia
35	162	14.8	349	21 AAG38205	Arabidopsis thalia
36	162	14.8	351	21 AAG38204	Arabidopsis thalia
37	160.5	14.7	181	21 AAG37809	Arabidopsis thalia
38	160.5	14.7	331	20 AAG3088	Human CPRM3 protei
39	159.5	14.6	330	20 AAW93086	Human CPRM1 protei
40	158	14.5	329	19 AAW71534	Helicobacter polyp
41	156.5	14.3	327	21 AAB09903	Rat aflatoxin B1 a
42	156.5	14.3	338	20 AAW93089	Mouse mCPRM1 allel
43	156.5	14.3	338	20 AAW93090	Mouse mCPRM1 allel
44	154.5	14.2	278	21 AAG36393	Arabidopsis thalia
45	154.5	14.2	286	21 AAG36392	Arabidopsis thalia

ALIGNMENTS

RESULT 1
AAW37716
ID AAW37716 standard; Protein; 216 AA.
AC AAW37716;
XX
XX
DT 12-MAR-1998 (first entry)
XX
DE C. glutamicum orf3 protein.
XX
KW LysG; LysE; ORF3; lysine transport; regulatory protein; export protein;
KW Microbial production; amino acid; animal feed additive.
XX
OS Corynebacterium glutamicum.
XX
PN DE19548222-A1.
XX
PD 26-JUN-1997.
XX
PF 22-DEC-1995; 95DE-1048222.
XX
PR 22-DEC-1995; 95DE-1048222.
XX
(KERJ) FORSCHUNGSZENTRUM JUELICH GMBH.
XX
PI Eggeling L, Sahm H, Vrljic M;
XX
DR WPI; 1997-333867/31.
DR N-PSDB; AAT96816.
XX
PT Increasing microbial production of amino acids, especially lysine -
PT by improving export carrier activity or corresponding gene
PT expression, also new export and regulatory genes from
PT Corynebacterium

```

XX Disclosure; Page 9; 16pp; German.
XX
CC This sequence is the Orf3 protein product, involved in lysine production.
CC LysG and LysE encode a lysine transport regulatory protein and an export
CC protein, respectively. Microbial production of amino acids (A) is
CC improved by increasing the export-carrier activity and/or the export gene
CC expression in a microorganism that produces (A). The method is
CC specifically used to increase production of lysine, used as an animal
CC feed additive. Other (A) are variously useful as pharmaceuticals,
CC condiments and intermediates for fine chemicals. This method increases
CC the amount of (A) secreted into the culture medium. Export of (A) has
CC been found to depend on a single gene.
XX Sequence      216 AA;
XX
    Query Match          100.0%; Score 1091; DB 18; Length 216;
    Best Local Similarity 100.0%; Pred. No. 1.3e-102;
    Matches 216; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY   1 DTPLEETMYALRDIVASGKALVYGISSVGPELTAAAEFMAEEGCPLLIHQPSYSILNRW 60
DB   1 dtpleetmyalrdivavgkalyvgissygpeltataaeafmaeegcpllihqpysilnrw 60
QY   61 VEEPGDGENLQSAANNGLGVIAFSPLAQGLLTDKYLDGIPEGSRASQGKLSBGMNLNV 120
DB   61 veepgdgenllgsaannglgiavfsplaqglldtkylldgipegrasqgkslsegmnlv 120
QY   121 NNIDMWKRLNDIAERGOSLAOMALAWYLREGQEGYADTVTSALIGASSVEQLDMSLDSL 180
DB   121 nnidmvrklndiaergogslaqmalawylreggegyadvtvsaligassveqldnslds 180
QY   181 NNLFFSDAELEAIDEISHDAGINIWAKATDSKTREN 216
DB   181 nnlffsdaealeideishdaginiwakatdsktren 216
XX RESULT      2
XX AAB76844
XX ID AAB76844 standard; Protein; 290 AA.
XX AC AAB76844;
XX XX
XX DT 11-APR-2001 (first entry)
XX DE
XX DE
XX KW Corynebacterium glutamicum MCT protein SEQ ID NO:670.
XX KW Corynebacterium glutamicum; brevivibacterium lactofermentum; MCT;
XX KW membrane construction and membrane transport protein; petroleum spill;
XX KW hydrocarbon degradation; gram positive aerobic bacterium; marker;
XX KW identification; microorganism; fine chemical production; transformation;
XX KW genome mapping; genetic engineering.
XX OS
XX OS Corynebacterium glutamicum.
XX PN W0200100805-A2.
XX XX
XX PD 04-JAN-2001.
XX PF
XX PF 23-JUN-2000; 2000WO-IB00926.
XX PR
XX PR 25-JUN-1999; 99US-0141031.
XX PR 08-JUL-1999; 99DE-1031454.
XX PR 08-JUL-1999; 99DE-1031478.
XX PR 08-JUL-1999; 99DE-1031563.
XX PR 09-JUL-1999; 99DE-1032122.
XX PR 09-JUL-1999; 99DE-1032124.
XX PR 09-JUL-1999; 99DE-1032125.
XX PR 09-JUL-1999; 99DE-1032128.
XX PR 09-JUL-1999; 99DE-1032180.
XX PR 09-JUL-1999; 99DE-1032182.
XX PR 09-JUL-1999; 99DE-1032190.

```

Db	156	ntpieetlealndvwtgkaryigasmshasqfaqalelqkqhgwaqfsmqdhynliyz	213
Qy	60	WVEEPGDGGENLLQSRANNGLGVIATFSPLAQGLLTQKYLDTGIEGSRASQ-----GKSL	113
Db	216	-----eeeremlpkyqegvavipwspplarglrtr-----pwgettarlvdsdevgknl	263
Qy	114	SEGMLNVNIDMVRKLNNDIAQERQSGSLAQMALAWLWIREQGEYGADVTVTSLICASSVEQL	173
Db	264	ykes-dendaqiaerlgtvseeelgatraqvalawllskpg-----laapligtsreeql	316
Qy	174	DNSLDLSLN 181	
Db	317	deilnavd 324	
RESULT 4			
AAW23388	ID	AAW23388 standard; protein; 329 AA.	
XX	AC		
XX	AAW23388;		
DT	DT		
XX	02-APR-1998	(first entry)	
DE	Shaker-like potassium ion channel beta-subunit core region Kv beta 1.		
XX	Shaker-like potassium ion channel; SPC; beta subunit; core region;		
KW	Kv beta 1; N-terminal A and B box; NAB; treatment; cardiac disease;		
KW	tumour; auto immune disease.		
XX	Mammalia.		
XX	WO9731112-A2.		
PN	28-AUG-1997.		
PD			
XX	18-FEB-1997; 97WO-US02292.		
PF	23-FEB-1996; 96US-0606143.		
XX	(UYJO) UNIV JOHNS HOPKINS SCHOOL MEDICINE.		
PA	Li M;		
PI	WPI; 1997-435164/40.		
XX			
XX	Polyptide(s) derived from Shaker-like potassium ion channel alpha		
PT	and beta subunits - used to alter potassium ion levels in a cell,		
PT	e.g. for treating neurological disorders, tumours, metabolic disease		
PT	and cardiac disease		
XX	Claim 15; Pages 51-52; 106pp; English.		
PS			
XX	This polypeptide sequence Kv beta 1 consists of the core region of a		
CC	beta-subunit of a Shaker-like potassium ion channel (SPC). This core		
CC	region can bind to a polypeptide consisting of the N-terminal A and B		
CC	box (NAB) domain and the NAB-S1 (the first transmembrane spanning dom		
CC	linking region of the alpha-subunit of SPC. The polypeptides or the		
CC	nucleic acid encoding them can be introduced into the cytoplasm of a		
CC	cell to modulate the flow of potassium ions through a cytoplasmic cel		
CC	membrane. Potassium ion channels regulate the action potentials, card		
CC	pacemaking and neurotransmitter release in excitable tissues. In		
CC	non-excitable tissues they play important roles in hormone secretion,		
CC	cell proliferation, cell volume regulation and lymphocyte		
CC	differentiation. Molecules which bind to the alpha or beta-subunit		
CC	polypeptides can be detected by contacting the polypeptides, with a		
CC	putative NAB and NAB-S1 linking region of an alpha-subunit or with a		
CC	putative core region of a beta-subunit respectively, and determining		
CC	whether or not binding occurs. These polypeptides and the encoding		
CC	nucleic acids may be useful in the treatment of a huge variety of		
CC	disorders, e.g. neurological disorders, tumours, metabolic diseases,		
CC	cardiac disease and autoimmune disease.		
XX	Sequence 329 AA;		
SQ			

CC		putative NAB and NAB-SI linking region of an alpha-subunit or with
CC		putative core region of a beta-subunit respectively, and determining
CC		whether or not binding occurs. These polypeptides and the encoding
CC		nucleic acids may be useful in the treatment of a huge variety of
CC		disorders, e.g. neurological disorders, tumours, metabolic diseases,
CC		cardiac disease and autoimmune disease.
XX		
SQ	Sequence	329 AA;
	Query Match	18.5%; Score 202; DB 18; Length 329;
	Best Local Similarity	31.2%; Pred. No. 3.1e-12;
	Matches	64; Conservative 37; Mismatches 82; Indels 22; Gaps
Qy	1 DTPLEETMYALRDIVASGKALYGVIGSYGPDLTAEAAEFMAEEG-CPLLTHQPYSYINR	59
Dd	: :: : : :: : :	:: : : : : : : :
Dd	125 ntpmeeivramthinvgmmanywqtswssmeimeaysvargfnlfpiceaqeyhmfrq	184
Qy	60 WVEEPGGDGENLQLSAANNGLVIAFSLAQQGLLTDRKYLDPGEGRSAS-----QGKS	112
Dd	: : : : : : : : : : : : : : :	:
Dd	185 ekve----vqlpel fhkigvgamt w3placg ivsgkyds gippysras lkg yqw lkdki	239
Qy	113 LSEGMNVNNDWVRKLNDTAERGOSLAQMALAWVLREGEYGADVTTSALIGASSVSEQ	172
Dd	: : : : : : : : : : : : : : : :	:
Dd	240 lsee--grrqakikeiqalaeirlgctlpqlatawcirneg-----vssvlignasnaeq	291
Qy	173 LDNSLDLSNNL-EFSDAELEAIDEI	196
Dd	: : : : : : : : : : : :	:
Dd	292 lmenigaiqvlpklsssivheidsi	316
RESULT	6	
AAG35920	ID	AAG35920 standard; Protein; 266 AA.
XX	AC	AAG35920;
XX	DT	
XX	DT	18-OCT-2000 (first entry)
XX	DE	Zea mays protein fragment SEQ ID NO: 43947.
KW	Protein identification; signal transduction pathway; metabolic pathway	
KW	hybridisation assay; genetic mapping; gene expression control; promotio	
KW	termination sequence; corn.	
XX		
OS	Zea mays subsp. mays.	
XX		
PN	EP1033405-A2.	
XX		
PD	06-SEP-2000.	
XX		
Pf	25-FEB-2000; 2000EP-0301439.	
XX		
PR	25-FEB-1999;	99US-0121825.
PR	05-MAR-1999;	99US-0123180.
PR	09-MAR-1999;	99US-0123548.
PR	23-MAR-1999;	99US-0125788.
PR	23-MAR-1999;	99US-0126284.
PR	29-MAR-1999;	99US-0126785.
PR	01-APR-1999;	99US-0127462.
PR	06-APR-1999;	99US-0128234.
PR	08-APR-1999;	99US-0128714.
PR	16-APR-1999;	99US-0129845.
PR	19-APR-1999;	99US-0130077.
PR	21-APR-1999;	99US-0130449.
PR	23-APR-1999;	99US-0130510.
PR	23-APR-1999;	99US-0130891.
PR	28-APR-1999;	99US-0131449.
PR	30-APR-1999;	99US-0132048.
PR	30-APR-1999;	99US-0132407.
PR	04-MAY-1999;	99US-0132484.
PR	05-MAY-1999;	99US-0132485.
PR	05-MAY-1999;	99US-0132486.

[illegible]

PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139753.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140895.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144632.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 22-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0132487.
PR 23-JUL-1999; 99US-0132863.
PR 23-JUL-1999; 99US-0134256.
PR 26-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149802.
PR 23-AUG-1999; 99US-0149803.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 01-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159394.
PR 13-OCT-1999; 99US-0159495.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 18-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.

PR	21-OCT-1999;	99US-0160741
PR	21-OCT-1999;	99US-0160767
PR	21-OCT-1999;	99US-0160768
PR	21-OCT-1999;	99US-0160770
PR	21-OCT-1999;	99US-0160814
PR	21-OCT-1999;	99US-0160815
PR	21-OCT-1999;	99US-0160980
PR	22-OCT-1999;	99US-0160981
PR	22-OCT-1999;	99US-0161049
PR	25-OCT-1999;	99US-0161404
PR	25-OCT-1999;	99US-0161405
PR	25-OCT-1999;	99US-0161406
PR	26-OCT-1999;	99US-0161359
PR	26-OCT-1999;	99US-0161360
PR	26-OCT-1999;	99US-0161361
PR	28-OCT-1999;	99US-0161920
PR	28-OCT-1999;	99US-0161992
PR	28-OCT-1999;	99US-0161993
PR	28-OCT-1999;	99US-0162142
PR	29-OCT-1999;	99US-0162142

[illegible]

REC'D
AAG09266
ID AAG09266 standard: Protein: 216 AA

17-OCT-2000 (first entry)

Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter
KW termination sequence.

PN EP1033405-A2

PF 25-FEB-2000; 2000EP-0301439.

PR 09-MAR-1999; 99US-0123548.

PR 01-APR-1999; 99US-0127462.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

PR 19-APR-1999; 99US-0130077.
 PR 21-APR-1999; 99US-0130449.
 PR 23-APR-1999; 99US-0130510.
 PR 23-APR-1999; 99US-0130891.
 PR 28-APR-1999; 99US-0131449.
 PR 30-APR-1999; 99US-0132048.
 PR 30-APR-1999; 99US-0132407.
 PR 04-MAY-1999; 99US-0132484.
 PR 05-MAY-1999; 99US-0132485.
 PR 06-MAY-1999; 99US-0132486.
 PR 06-MAY-1999; 99US-0132487.
 PR 07-MAY-1999; 99US-0132863.
 PR 11-MAY-1999; 99US-0132856.
 PR 14-MAY-1999; 99US-0134218.
 PR 14-MAY-1999; 99US-0134219.
 PR 14-MAY-1999; 99US-0134221.
 PR 14-MAY-1999; 99US-0134370.
 PR 18-MAY-1999; 99US-0134768.
 PR 19-MAY-1999; 99US-0134941.
 PR 20-MAY-1999; 99US-0135124.
 PR 21-MAY-1999; 99US-0135353.
 PR 24-MAY-1999; 99US-0135629.
 PR 25-MAY-1999; 99US-0136021.
 PR 27-MAY-1999; 99US-0136392.
 PR 28-MAY-1999; 99US-0136782.
 PR 01-JUN-1999; 99US-0137222.
 PR 03-JUN-1999; 99US-0137528.
 PR 04-JUN-1999; 99US-0137502.
 PR 07-JUN-1999; 99US-0137724.
 PR 08-JUN-1999; 99US-0138094.
 PR 10-JUN-1999; 99US-0138540.
 PR 10-JUN-1999; 99US-0138847.
 PR 14-JUN-1999; 99US-0139119.
 PR 16-JUN-1999; 99US-0139452.
 PR 16-JUN-1999; 99US-0139453.
 PR 17-JUN-1999; 99US-0139492.
 PR 18-JUN-1999; 99US-0139454.
 PR 18-JUN-1999; 99US-0139455.
 PR 18-JUN-1999; 99US-0139456.
 PR 18-JUN-1999; 99US-0139457.
 PR 18-JUN-1999; 99US-0139458.
 PR 18-JUN-1999; 99US-0139459.
 PR 18-JUN-1999; 99US-0139460.
 PR 18-JUN-1999; 99US-0139461.
 PR 18-JUN-1999; 99US-0139462.
 PR 18-JUN-1999; 99US-0139463.
 PR 18-JUN-1999; 99US-0139750.
 PR 18-JUN-1999; 99US-0139763.
 PR 21-JUN-1999; 99US-0139817.
 PR 22-JUN-1999; 99US-0139899.
 PR 23-JUN-1999; 99US-0140353.
 PR 23-JUN-1999; 99US-0140354.
 PR 24-JUN-1999; 99US-0140695.
 PR 28-JUN-1999; 99US-0140823.
 PR 29-JUN-1999; 99US-0140991.
 PR 30-JUN-1999; 99US-0141287.
 PR 01-JUL-1999; 99US-0141842.
 PR 01-JUL-1999; 99US-0142154.
 PR 02-JUL-1999; 99US-0142055.
 PR 06-JUL-1999; 99US-0142390.
 PR 08-JUL-1999; 99US-0142803.
 PR 09-JUL-1999; 99US-0142920.
 PR 12-JUL-1999; 99US-0142977.
 PR 13-JUL-1999; 99US-0143542.
 PR 14-JUL-1999; 99US-0143624.
 PR 15-JUL-1999; 99US-0144005.
 PR 16-JUL-1999; 99US-0144085.
 PR 16-JUL-1999; 99US-0144086.
 PR 19-JUL-1999; 99US-0144332.
 PR 19-JUL-1999; 99US-0144331.
 PR 19-JUL-1999; 99US-0144332.
 PR 19-JUL-1999; 99US-0144333.
 PR 19-JUL-1999; 99US-0144334.

PR 19-JUL-1999; 99US-0144335.
 PR 20-JUL-1999; 99US-0144352.
 PR 20-JUL-1999; 99US-0144632.
 PR 20-JUL-1999; 99US-0144884.
 PR 21-JUL-1999; 99US-0144814.
 PR 21-JUL-1999; 99US-0145086.
 PR 21-JUL-1999; 99US-0145088.
 PR 22-JUL-1999; 99US-0145085.
 PR 22-JUL-1999; 99US-0145087.
 PR 22-JUL-1999; 99US-0145089.
 PR 22-JUL-1999; 99US-0145192.
 PR 23-JUL-1999; 99US-0145145.
 PR 23-JUL-1999; 99US-0145218.
 PR 23-JUL-1999; 99US-0145224.
 PR 26-JUL-1999; 99US-0145276.
 PR 27-JUL-1999; 99US-0145913.
 PR 27-JUL-1999; 99US-0145918.
 PR 27-JUL-1999; 99US-0145919.
 PR 28-JUL-1999; 99US-0145951.
 PR 02-AUG-1999; 99US-0146386.
 PR 02-AUG-1999; 99US-0146388.
 PR 02-AUG-1999; 99US-0146389.
 PR 03-AUG-1999; 99US-0147038.
 PR 04-AUG-1999; 99US-0147204.
 PR 04-AUG-1999; 99US-0147302.
 PR 05-AUG-1999; 99US-0147192.
 PR 06-AUG-1999; 99US-0147260.
 PR 06-AUG-1999; 99US-0147303.
 PR 06-AUG-1999; 99US-0147416.
 PR 09-AUG-1999; 99US-0147493.
 PR 09-AUG-1999; 99US-0147935.
 PR 10-AUG-1999; 99US-0148171.
 PR 11-AUG-1999; 99US-0148319.
 PR 12-AUG-1999; 99US-0148341.
 PR 13-AUG-1999; 99US-0148565.
 PR 13-AUG-1999; 99US-0148684.
 PR 16-AUG-1999; 99US-0149368.
 PR 17-AUG-1999; 99US-0149175.
 PR 18-AUG-1999; 99US-0149426.
 PR 20-AUG-1999; 99US-0149722.
 PR 20-AUG-1999; 99US-0149723.
 PR 20-AUG-1999; 99US-0149929.
 PR 23-AUG-1999; 99US-0149902.
 PR 23-AUG-1999; 99US-0149930.
 PR 25-AUG-1999; 99US-0150566.
 PR 26-AUG-1999; 99US-0150884.
 PR 27-AUG-1999; 99US-0151065.
 PR 27-AUG-1999; 99US-0151066.
 PR 27-AUG-1999; 99US-0151080.
 PR 30-AUG-1999; 99US-0151303.
 PR 31-AUG-1999; 99US-0151438.
 PR 01-SEP-1999; 99US-0151930.
 PR 07-SEP-1999; 99US-0152363.
 PR 10-SEP-1999; 99US-0153070.
 PR 13-SEP-1999; 99US-0153758.
 PR 15-SEP-1999; 99US-0154018.
 PR 16-SEP-1999; 99US-0154039.
 PR 20-SEP-1999; 99US-0154779.
 PR 22-SEP-1999; 99US-0155139.
 PR 23-SEP-1999; 99US-0155486.
 PR 24-SEP-1999; 99US-0155659.
 PR 28-SEP-1999; 99US-0156458.
 PR 29-SEP-1999; 99US-0156596.
 PR 04-OCT-1999; 99US-0157117.
 PR 05-OCT-1999; 99US-0157753.
 PR 06-OCT-1999; 99US-0157865.
 PR 07-OCT-1999; 99US-0158029.
 PR 08-OCT-1999; 99US-0158232.
 PR 12-OCT-1999; 99US-0158369.
 PR 13-OCT-1999; 99US-0159293.
 PR 13-OCT-1999; 99US-0159294.
 PR 13-OCT-1999; 99US-0159295.
 PR 14-OCT-1999; 99US-0159329.

Query Match	17.98;	Score 195;	DB 21;	Length 216;
Best Local Similarity	29.08;	Pred. No. 8.5e-12;		
Matches 61;	Conservative 41;	Mismatches 72;	Indels 36;	Gaps 9;
PR 14-OCT-1999;	99US-0159330.			
PR 14-OCT-1999;	99US-0159331.			
PR 14-OCT-1999;	99US-0159637.			
PR 14-OCT-1999;	99US-0159638.			
PR 18-OCT-1999;	99US-0159584.			
PR 21-OCT-1999;	99US-0160741.			
PR 21-OCT-1999;	99US-0160767.			
PR 21-OCT-1999;	99US-0160768.			
PR 21-OCT-1999;	99US-0160770.			
PR 21-OCT-1999;	99US-0160814.			
PR 21-OCT-1999;	99US-0160815.			
PR 22-OCT-1999;	99US-0160980.			
PR 22-OCT-1999;	99US-0160981.			
PR 22-OCT-1999;	99US-0160989.			
PR 25-OCT-1999;	99US-0161404.			
PR 25-OCT-1999;	99US-0161405.			
PR 25-OCT-1999;	99US-0161406.			
PR 26-OCT-1999;	99US-0161359.			
PR 26-OCT-1999;	99US-0161360.			
PR 26-OCT-1999;	99US-0161361.			
PR 28-OCT-1999;	99US-0161920.			
PR 28-OCT-1999;	99US-0161921.			
PR 28-OCT-1999;	99US-0161931.			
PR 29-OCT-1999;	99US-0162142.			
QY 2	TPLEETMYALRDIVASGKALVIGISSYGPETATAAEFMAEEGCP---	LIHQPSYSIN 58		
Db 16	tpleetramyvvdkkqwfayvgwtsewaaqitea--wgaardrlvlgpiveqeynmfa 73			
QY 59	RWVEEPGDGDLNLSAANGLGVIATFSPLAGGLITDKYLDG--IPEGSRASQKSLSEG 117			
Db 74	rhkve-----teflpyncnhgigttsplagsvltgkynkalpsdsrfa-----i 120			
QY 118	LNVANI-----DMVRK---LNDIAQERQSGLAQMALAWLREQEGYGDVTTSALICA 167			
Db 121	enynknanrslvddvrlkvsqkpiadelgtvltlaqlaiaawcasn-----pnvssvitga 174			
QY 168	SSVQLDNLSDLSANLE-FSDAELEAIDEI 196			
Db 175	tresqigenmkavdviplltpivldklegv 204			
RESULT 9				
ID AAG39051	standard; Protein; 216 AA.			
XX AC AAG39051;				
XX DT 18-OCT-2000	(first entry)			
XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 48265.				
XX KW Protein identification; signal transduction pathway; metabolic pathway;				
XX KW hybridisation assay; genetic mapping; gene expression control; promoter;				
XX KW termination sequence.				
XX OS Arabidopsis thaliana.				
XX PN EP1033405-A2.				
XX PD 06-SEP-2000.				
XX XX 25-FEB-2000;	2000EP-0301439.			
PR 25-FEB-1999;	99US-0121825.			
PR 05-MAR-1999;	99US-0123180.			
PR 09-MAR-1999;	99US-0123548.			
PR 23-MAR-1999;	99US-0125788.			
PR 25-MAR-1999;	99US-0126264.			
PR 29-MAR-1999;	99US-0126789.			
PR 01-APR-1999;	99US-0127462.			
PR 06-APR-1999;	99US-0128234.			
PR 08-APR-1999;	99US-0128714.			
PR 16-APR-1999;	99US-0129845.			
PR 19-APR-1999;	99US-0130077.			
PR 21-APR-1999;	99US-0130449.			
PR 23-APR-1999;	99US-0130510.			
PR 23-APR-1999;	99US-0130891.			
PR 28-APR-1999;	99US-0131449.			
PR 30-APR-1999;	99US-0132048.			
PR 30-APR-1999;	99US-0132407.			
PR 04-MAY-1999;	99US-0132484.			
PR 05-MAY-1999;	99US-0132485.			
PR 06-MAY-1999;	99US-0132486.			
PR 07-MAY-1999;	99US-0132487.			
PR 07-MAY-1999;	99US-0132863.			
PR 11-MAY-1999;	99US-0134256.			
PR 14-MAY-1999;	99US-0134218.			
PR 14-MAY-1999;	99US-0134219.			
PR 14-MAY-1999;	99US-0134221.			

PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-01322407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.

PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 22-SEP-1999; 99US-0154779.
PR 23-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.

PR	04-OCT-1999;	99US-0157117.	XX	06-SEP-2000.	
PR	05-OCT-1999;	99US-0157753.	PD		
PR	06-OCT-1999;	99US-0157865.	XX		
PR	07-OCT-1999;	99US-0158029.	PF	25-FEB-2000;	2000EP-0301439.
PR	08-OCT-1999;	99US-0158232.	XX		
PR	12-OCT-1999;	99US-0158369.	PR	25-FEB-1999;	99US-0121825.
PR	13-OCT-1999;	99US-0159293.	PR	05-MAR-1999;	99US-0123180.
PR	13-OCT-1999;	99US-0159294.	PR	09-MAR-1999;	99US-0123548.
PR	13-OCT-1999;	99US-0159295.	PR	23-MAR-1999;	99US-0125788.
PR	14-OCT-1999;	99US-0159329.	PR	25-MAR-1999;	99US-0126264.
PR	14-OCT-1999;	99US-0159330.	PR	29-MAR-1999;	99US-0126785.
PR	14-OCT-1999;	99US-0159331.	PR	01-APR-1999;	99US-0127462.
PR	14-OCT-1999;	99US-0159637.	PR	06-APR-1999;	99US-0128234.
PR	14-OCT-1999;	99US-0159638.	PR	08-APR-1999;	99US-0128714.
PR	18-OCT-1999;	99US-0159584.	PR	16-APR-1999;	99US-0129845.
PR	21-OCT-1999;	99US-0160741.	PR	19-APR-1999;	99US-0130077.
PR	21-OCT-1999;	99US-0160767.	PR	21-APR-1999;	99US-0130449.
PR	21-OCT-1999;	99US-0160768.	PR	23-APR-1999;	99US-0130510.
PR	21-OCT-1999;	99US-0160770.	PR	23-APR-1999;	99US-0130891.
PR	21-OCT-1999;	99US-0160814.	PR	28-APR-1999;	99US-0131449.
PR	21-OCT-1999;	99US-0160815.	PR	30-APR-1999;	99US-0132048.
PR	22-OCT-1999;	99US-0160980.	PR	30-APR-1999;	99US-0132407.
PR	22-OCT-1999;	99US-0160981.	PR	04-MAY-1999;	99US-0132484.
PR	22-OCT-1999;	99US-0160989.	PR	05-MAY-1999;	99US-0132485.
PR	25-OCT-1999;	99US-0161404.	PR	06-MAY-1999;	99US-0132486.
PR	25-OCT-1999;	99US-0161405.	PR	06-MAY-1999;	99US-0132487.
PR	25-OCT-1999;	99US-0161406.	PR	07-MAY-1999;	99US-0132863.
PR	26-OCT-1999;	99US-0161359.	PR	11-MAY-1999;	99US-0134256.
PR	26-OCT-1999;	99US-0161360.	PR	14-MAY-1999;	99US-0134218.
PR	26-OCT-1999;	99US-0161361.	PR	14-MAY-1999;	99US-0134219.
PR	28-OCT-1999;	99US-0161920.	PR	14-MAY-1999;	99US-0134221.
PR	28-OCT-1999;	99US-0161992.	PR	18-MAY-1999;	99US-0134370.
PR	28-OCT-1999;	99US-0161993.	PR	18-MAY-1999;	99US-0134768.
PR	29-OCT-1999;	99US-0162142.	PR	19-MAY-1999;	99US-0134941.
PR			PR	20-MAY-1999;	99US-0135124.
PR			PR	21-MAY-1999;	99US-0135353.
PR			PR	24-MAY-1999;	99US-0135629.
PR			PR	25-MAY-1999;	99US-0136021.
PR			PR	27-MAY-1999;	99US-0136392.
PR			PR	28-MAY-1999;	99US-0136782.
PR			PR	01-JUN-1999;	99US-0137222.
PR			PR	03-JUN-1999;	99US-0137528.
PR			PR	04-JUN-1999;	99US-0137502.
PR			PR	07-JUN-1999;	99US-0137724.
PR			PR	08-JUN-1999;	99US-0138094.
PR			PR	10-JUN-1999;	99US-0138540.
PR			PR	10-JUN-1999;	99US-0138847.
PR			PR	14-JUN-1999;	99US-0139119.
PR			PR	16-JUN-1999;	99US-0139452.
PR			PR	16-JUN-1999;	99US-0139453.
PR			PR	17-JUN-1999;	99US-0139454.
PR			PR	18-JUN-1999;	99US-0139454.
PR			PR	18-JUN-1999;	99US-0139455.
PR			PR	18-JUN-1999;	99US-0139456.
PR			PR	18-JUN-1999;	99US-0139457.
PR			PR	18-JUN-1999;	99US-0139458.
PR			PR	18-JUN-1999;	99US-0139459.
PR			PR	18-JUN-1999;	99US-0139460.
PR			PR	18-JUN-1999;	99US-0139461.
PR			PR	18-JUN-1999;	99US-0139462.
PR			PR	18-JUN-1999;	99US-0139463.
PR			PR	18-JUN-1999;	99US-0139750.
PR			PR	18-JUN-1999;	99US-0139763.
PR			PR	21-JUN-1999;	99US-0139817.
PR			PR	22-JUN-1999;	99US-0139899.
PR			PR	23-JUN-1999;	99US-0140353.
PR			PR	23-JUN-1999;	99US-0140354.
PR			PR	24-JUN-1999;	99US-0140695.
PR			PR	28-JUN-1999;	99US-0140823.
PR			PR	29-JUN-1999;	99US-0140991.
PR			PR	30-JUN-1999;	99US-0141287.
PR			PR	01-JUL-1999;	99US-0141842.
PR			PR	01-JUL-1999;	99US-0142154.

Query Match 17.9%; Score 195; DB 21; Length 266;
 Best Local Similarity 29.0%; Pred. No. 1.2e-11;
 Matches 61; Conservative 41; Mismatches 72; Indels 36; Gaps 9;

QY 2 TPLEETMYALRDIVASGKALYGISSYGPGLTAEAFMAEECCPL---LIHOPSYSIIN 58
 Db 66 tpleetvramnyvidkgwafygtsewsaqqitea--wgaadrldvgpiveqpeynmfa 123

QY 59 RWVEEPDGDGNLLQSAANNGLGVAFSPLAQGLLTDKYLDG-IPGSRASQCKSLSEGM 117
 Db 124 rhkve-----teflplythnhglttwsplagvltgkynkaipsdsrfa-----l 170

QY 118 LNVNNT-----DMVRK---LNDIAQERGSLAQMALAWLREQGEYGADTVTSALIGA 167
 Db 171 enyknlanrslvddvirkvsgikpiadelgvtlaqlaiaawcasn-----pnvssvitga 224

QY 168 SSVEQLDNLSDLNLE-FSDAELELAIDEI 196
 Db 225 tresqigenmkadvipllitpivldkieqv 254

RESULT 11
 AAG39049
 ID AAG39049 standard; Protein; 328 AA.
 AC AAG39049;
 XX
 DT 18-OCT-2000 (first entry)
 XX
 DE Arabidopsis thaliana protein fragment SEQ ID NO: 48263.
 XX
 KW Protein identification; signal transduction pathway; metabolic pathway;
 KW hybridisation assay; genetic mapping; gene expression control; promoter;
 KW termination sequence.
 XX
 OS Arabidopsis thaliana.
 XX
 PN EP1033405-A2.

fine chemical production; organic acid; proteinogenic amino acid; nonproteinogenic amino acid; purine base; pyrimidine base; nucleoside; nucleotide; lipid; saturated fatty acid; unsaturated fatty acid; diol; carbohydrate; aromatic compound; vitamin; cofactor; polyketide; enzyme; diagnosis; *Corynebacterium diphtheriae*; evolutionary study.

Corynebacterium glutamicum.

WO200100844-A2.

04-JAN-2001.

23-JUN-2000; 2000WO-IB00943.

25-JUN-1999: 99US-0141031

08-JUL-1999; 99DE-1031412;

08-JUL-1999; 99DE-1031413.

R 08-JUL-1999; 99DE-1031419.

08-JUL-1999; 99DE-1031420.

08-JUL-1999; 99DE-1031424.

08-JUL-1999; 99DE-1031428.
08-JUL-1999; 99DE-1031431

08-JUL-1999; 99DE-1031431.
08-JUL-1999; 99DE-1031433

08-JUL-1999; 99DE-1031434.

08-JUL-1999; 99DE-1031510.

08-JUL-1999; 99DE-1031562.

08-JUL-1999; 99DE-1031634.

09-JUL-1999; 99DE-1032180.

09-JUL-1999;
09-JUL-1999;
09-JUL-1999;

09-JUL-1999: 99JIS-0143208
09-JUL-1999: 99JIS-0143208

14-JUL-1999; 99DE-1032924;

14-JUL-1999; 99DE-1032973.

R 14-JUL-1999; 99DE-1033005.

27-AUG-1999; 99DE-1040765.

31-AUG-1999; 99US-0151572.

03-SEP-1999; 99DE-1042076.
03-SEP-1999; 99DE-1042070

03-SEP-1999; 99DE-1042079.
03-SEP-1999; 99DE-1042086

03-SEP-1999: 99DE-1042087:

03-SEP-1999; 99DE-1042088.

03-SEP-1999; 99DE-1042095.

03-SEP-1999; 99DE-1042123.

03-SEP-1999; 99DE-1042125.

/PA DT \ PAGE AC

(BAU) / BASE AG.

Pompeius M. Kro

2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841

N-PSDB; AAF71735.

metabolism and oxidative phosphorylation protein for production or modulation of production of fine chemicals e.g. amino acids, carbohydrates or enzymes -

Claim 20; Page 1208-1209; 1246pp; English.

AAAF1360 to AAAF1750 encode the *Corynebacterium glutamicum* sugar metabolism and oxidative phosphorylation (SMP) proteins given in AAAF9243 to AAB 79633 which are involved in carbon metabolism and energy production. The *C. glutamicum* SMP gene can be used in vectors (II) for expression in host cells and production or modulation of production of fine chemicals, such as, an organic acid, a proteinogenic or nonproteinogenic amino acid (preferred), a purine or pyrimidine base, a nucleoside, a nucleotide, a lipid, a saturated or unsaturated fatty acid, a diol, a carbohydrate, an aromatic compound, a vitamin, a cofactor, a polyketide, or an enzyme. The presence of (I) or SMP proteins (III) encoded by them are used for diagnosing the presence or activity of *Corynebacterium diphtheriae* in a subject. (I), (II), (III) or host cells containing them are used to map genomes of organisms related to *C. glutamicum*, identify and localise *C. glutamicum* sequences of interest.

Db	184	rrae-----qellpatqhlvgvffagaplgggvltakyrseiphdsraastgrdaevqs	237
Qy	117	MLNVNNTDMVRKRLNDIAOERGQSLAQMALAWLREOGEYGADTWTSAIGASSVEQLDNS	176
Db	238	ylndrgriiwdaltdaakglgispavtattwvdrpg-----vtavivgartheqlshl	291
Qy	177	LDSLNNLEFSDAELEAIDEIS	197
Db	292	lka-esvtlptptqalddvs	311
RESULT 14			
AAB76845			
XX	ID	AAB76845 standard; Protein; 315 AA.	
XX	AC	AAB76845;	
XX	DT	11-APR-2001 (first entry)	
XX	DE	Corynebacterium glutamicum MCT protein SEQ ID NO:672.	
XX	KW	Corynebacterium glutamicum; brevibacterium lactofermentum; MCT;	
KW	KW	membrane construction and membrane transport protein; petroleum spill;	
KW	KW	hydrocarbon degradation; gram positive aerobic bacterium; marker;	
KW	KW	identification; microorganism; fine chemical production; transformation;	
KW	KW	genome mapping; genetic engineering.	
XX	OS	Corynebacterium glutamicum.	
XX	OS		
PN	WO	20010805-A2.	
XX	XX		
PD	04	-JAN-2001.	
XX	XX		
PF	23	-JUN-2000; 2000WO-IB00926.	
XX	XX		
PR	25	-JUN-1999; 99US-0141031.	
PR	08	-JUL-1999; 99DE-1031454.	
PR	08	-JUL-1999; 99DE-1031478.	
PR	08	-JUL-1999; 99DE-1031563.	
PR	09	-JUL-1999; 99DE-1032122.	
PR	09	-JUL-1999; 99DE-1032124.	
PR	09	-JUL-1999; 99DE-1032125.	
PR	09	-JUL-1999; 99DE-1032128.	
PR	09	-JUL-1999; 99DE-1032180.	
PR	09	-JUL-1999; 99DE-1032182.	
PR	09	-JUL-1999; 99DE-1032190.	
PR	09	-JUL-1999; 99DE-1032191.	
PR	09	-JUL-1999; 99DE-1032209.	
PR	09	-JUL-1999; 99DE-1032212.	
PR	09	-JUL-1999; 99DE-1032227.	
PR	09	-JUL-1999; 99DE-1032228.	
PR	09	-JUL-1999; 99DE-1032229.	
PR	09	-JUL-1999; 99DE-1032230.	
PR	14	-JUL-1999; 99DE-1032927.	
PR	14	-JUL-1999; 99DE-1033005.	
PR	14	-JUL-1999; 99DE-1033006.	
PR	27	-AUG-1999; 99DE-1040764.	
PR	27	-AUG-1999; 99DE-1040765.	
PR	27	-AUG-1999; 99DE-1040766.	
PR	27	-AUG-1999; 99DE-1040830.	
PR	27	-AUG-1999; 99DE-1040831.	
PR	27	-AUG-1999; 99DE-1040832.	
PR	27	-AUG-1999; 99DE-1040833.	
PR	31	-AUG-1999; 99DE-1041378.	
PR	31	-AUG-1999; 99DE-1041379.	
PR	31	-AUG-1999; 99DE-1041395.	
PR	03	-SEP-1999; 99DE-1042077.	
PR	03	-SEP-1999; 99DE-1042078.	
PR	03	-SEP-1999; 99DE-1042079.	
XX	XX	99DE-1042088.	
(BADI) BASF AG.			


```
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149375.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
```

```
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 16.5%; Score 180.5; DB 21; Length 316;
Best Local Similarity 26.3%; Pred. No. 4.4e-10;
Matches 52; Conservative 45; Mismatches 76; Indels 25; Gaps 6;

QY 3 PLEETMY--ALRDIVASGKALYGVGSISSYGPFLTAFAAEFMAEECCPLLIIHQPSYIINRW 60
Db 123 plqeralwdglvemyekglvsavgsnygpkqlkikhsylasrgvplssaqvqfslism- 181
QY 61 VEEPGDDGENLQSAANNGLVIAFSPLAOGLLTDKY-LDGIPEGSRASOGKSLSEGMLN 119
Db 182 ----gddqmelksvcdslgrrilaysplgmltgkygasnlpkgrsvifrrqlpl-- 235
QY 120 VNNIDMYVRKLNIDIAQERGQSLAQMALAWLVREQGEYGADTVTSALIGASSVEQLDNSLDS 179
Db 236 ---epllscrlrrlaekkgkmpqvainwcmckg-----tvpipgyktvrhwqenlga 284
QY 180 ----LNNLEFSDAELEAI 193
Db 285 lgwrlssaeeiselesaaam 302
```

Search completed: September 13, 2001, 17:51:44
Job time: 1115 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 13, 2001, 17:51:50 ; Search time 37.95 Seconds
(without alignments)
753.040 Million cell updates/sec

Title: US-09-105-117i-2

Perfect score: 1091

Sequence: 1 DTPLEETWYALLRDIVASGKA.....SHDAGINIWAKATDSKTREN 216

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 425026 seqs, 132305027 residues

Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL16:*
1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mhc.*
8: sp_organelle.*
9: sp_phase.*
10: sp_plant.*
11: sp_rodent.*
12: sp_unclassified.*
13: sp_vertebrate.*
14: sp_virus.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1054	96.6	209	2 P94634	P94634 corynebacte
2	529.5	48.5	346	2 Q46851	Q46851 escherichia
3	529	48.5	348	2 Q9EWJ0	Q9EWJ0 streptomyc
4	449	41.2	319	2 Q9X7S9	Q9X7S9 streptomyc
5	438.5	40.2	334	2 Q9ZB20	Q9ZB20 lactococcus
6	283.5	26.0	315	2 Q9RS10	Q9RS10 deinococcus
7	267	24.5	319	2 Q9WYE9	Q9WYE9 thermotoga
8	257	23.6	336	1 Q34185	Q34185 halobacteri
9	238	21.8	336	1 Q9H0U7	Q9H0U7 halobacteri
10	237.5	21.8	387	2 Q9PGD5	Q9PGD5 xylella fas
11	227.5	20.9	344	3 Q59826	Q59826 schizosacch
12	213	19.5	354	2 Q44328	Q44328 agrobacteri
13	211	19.3	345	2 Q9HWY1	Q9HWY1 pseudomonas
14	210.5	19.3	330	2 Q9FZ25	Q9FZ25 streptomyc
15	210.5	19.3	336	2 Q9LIS0	Q9LIS0 streptomyc
16	209	19.2	307	2 Q9RJCA	Q9RJCA streptomyc
17	209	19.2	330	10 Q82064	Q82064 solanum tub
18	209	19.2	367	6 Q27955	Q27955 bos taurus
19	209	19.2	367	6 Q9NOV9	Q9NOV9 oryctolagus

ALIGNMENTS

RESULT 1

P94634 ID P94634 PRELIMINARY; PRT; 209 AA.

AC P94634;

DT 01-MAY-1997 (TREMREL. 03, Created)

DT 01-MAY-1997 (TREMREL. 03, Last sequence update)

DT 01-MAR-2001 (TREMREL. 16, Last annotation update)

DE LYSE AND LYSO GENES.

OS Corynebacterium glutamicum (Brevibacterium flavum).

OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;

OC Actinomycetales; Corynebacterineae; Corynebacteriaceae;

OC Corynebacterium.

OC NCBI_taxid=1718;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=R127;

RX MEDLINE=97126810; PubMed=8971704;

RA Viljic M., Sahm H., Eggeling L.;

RT "A new type of transporter with a new type of cellular function: L-

lysine export from Corynebacterium glutamicum.";

RL Mol. Microbiol. 22:815-826(1996).

DR EMBL: X96471; CAA65325.1;

SQ SEQUENCE 209 AA; 22450 MW; 06549D44F0BC0100 CRC64;

Query Match 96.6%; Score 1054; DB 2; Length 209;

Best Local Similarity 100.0%; Pred. No. 1.2e-75;

Matches 209; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 MYALRDIVASGKALYGVIGSYSGELTAEAEFFMAEECCPLLHOPSYSIINRWVEEPD 67

DB 1 MYALRDIVASGKALYGVIGSYSGELTAEAEFFMAEECCPLLHOPSYSIINRWVEEPD 60

QY 68 GENLQSAANGLGVIAFSPLAQGLTDTKYLDTGIPESRASQKSLSEGLMNNVNDMVR 127

DB 61 GENLQSAANGLGVIAFSPLAQGLTDTKYLDTGIPESRASQKSLSEGLMNNVNDMVR 120

QY 128 KLNDIAQERGQSQAOMALAWLWREQGEYGDVTYSALIGASSVEQLDNLNLEFSD 187

DB 121 KLNDIAQERGQSQAOMALAWLWREQGEYGDVTYSALIGASSVEQLDNLNLEFSD 180

QY 188 AELEAIDEISHDAGINIWAKATDSKTREN 216

[illegible][illegible]

RESULT	7
Q9WYE9	
ID	Q9WYE9
AC	PRELIMINARY;
DT	Q9WYE9;
DT	01-NOV-1999 (TrEMBLrel. 12, Created)
DT	01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DE	01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE	K+ CHANNEL, BETA SUBUNIT.
GN	TM0313.
OS	Thermotoga maritima.
OC	Bacteria; Thermotogales; Thermotoga.

OX NCBI_TaxID=2336;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MSB8 / DSM 3109;
 RX MEDLINE=99287316; PubMed=10360571;
 RA Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,
 RA Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
 RA McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,
 RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
 RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
 RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
 RT "Evidence for lateral gene transfer between Archaea and Bacteria from
 RL Nature 399:323-329(1999)."
 DR EMBL: AE001713; AAD35401.1; -;
 DR TIGR: TW0313; -;
 DR InterPro: IPR001395; -;
 DR PRINTS: PRO0069; ALDRETRDTASE.
 KW Ionic channel.
 SQ SEQUENCE 319 AA; 36171 MW; E4B1F32F41719B39 CRC64;

Query Match 24.5%; Score 267; DB 2; Length 319;
 Best Local Similarity 33.8%; Pred. No. 2.4e-13;
 Matches 71; Conservative 43; Mismatches 70; Indels 26; Gaps 7;
 QY 1 DTPLEETMYALRDIVASGKALYVGISSYGPGLTAEAEFMAEERG-CPLLIHOPSYSIINR 59
 DB 125 NVPMEEVVFAMDYILREGALVWGTSEWAKETEEAHRVCKELGVMPPIVEQPYNNFVR 184
 QY 60 WBEPEGDDGNNLQSA-----NNGLGVIASFPLAQGLLTKYLDGIPEGSRASQ-----GK 111
 DB 185 -----ERVEKEVAPLYEKYMGLTTPYSLASGLSKYNNNGIPEGSRRLATFPQVRK 235
 QY 112 SLSE-CMLNVNNDIMVRKLNDAIQAERGQSLAQMALAWLREGEYGDVTSALIGASSV 170
 DB 236 WLEEGLLNEKTFKRLKQNTAQGLSPLQALIAWILNK-----NVSSVILGVSRP 289
 QY 171 EQLDNLSLNNLE-FSDALELAIDRSHD 199
 DB 290 EQLEENLKAIVEKEKTEDVMEEIEKILNE 319

RESULT 8
 O34185 PRELIMINARY; PRT; 336 AA.
 AC O34185;
 DT 01-JAN-1998 (TremBLrel. 05, Created)
 DT 01-JAN-1998 (TremBLrel. 05, Last sequence update)
 DE OXIDOREDUCTASE.
 OS Halobacterium volcanii (Haloflex volcanii).
 OG Plasmid pV3.
 OC Archaea; Euryarchaeota; Halobacteriales; Halobacteriaceae; Haloflex.
 OX NCBI_TaxID=2246;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=DS2;
 RA Farghani R., Imbeault J.-C., St Jean A., Chan C.C.-Y., Allard G.,
 RA Charlebois R.L.;
 RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL: U95374; AAB71807.1; -;
 KW Plasmid.
 SQ SEQUENCE 336 AA; 37976 MW; 4CD19CE153A5B5 CRC64;

Query Match 23.6%; Score 257; DB 1; Length 336;
 Best Local Similarity 32.1%; Pred. No. 1.6e-12;
 Matches 67; Conservative 40; Mismatches 84; Indels 18; Gaps 7;
 QY 1 DTPLEETMYALRDIVASGKALYVGISSYGPGLTAEAEFMAEERG-CPLLIHOPSYSIINR 57
 DB 141 ETPTEMKTLNGLVEDGKLVHGLGASTLRPNKAWKARAEIARAEWGEPFVAQPRYNLV 200

QY 58 NRWVEEPGDDGNNLQSAANGLGVIAFSPLAQGLLTKYL--DGIPGSRASQKSLSE 115
 DB 201 DREIE--GD-----YLEMTRSYGIACVPCWPLGQGFUGTGTREDGUTGESRAESRFEE 254
 QY 116 GMLNVNNDIMVRKLNDAIQAERGQSLAQMALAWLREGEYGDVTSALIGASSVBOLEDN 175
 DB 255 SYLTENFVHDELDAVEVDATPAQTALAWLHRDG-----VTAPIVGARTVQLETFE 308
 QY 176 SLDSLNNLEFSDAELEAIDEISHDAGINI 204
 DB 309 NLEAA-TIDLTDEQVDRLTCAKPDYVGL 336
 RESULT 9
 Q9HQU7 PRELIMINARY; PRT; 336 AA.
 AC Q9HQU7;
 DT 01-MAR-2001 (TremBLrel. 16, Created)
 DT 01-MAR-2001 (TremBLrel. 16, Last sequence update)
 DE PROBABLE OXIDOREDUCTASE.
 GN YAJ02 OR VNG0998G.
 OS Halobacterium sp. (strain NRC-1).
 OC Archaea; Euryarchaeota; Halobacteriales; Halobacteriaceae;
 OC Halobacterium.
 OX NCBI_TaxID=64091;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20504483; PubMed=11016950;
 RA Ng W.V., Kennedy S.P., Mahairas G.G., Berquist B., Pan M.,
 RA Shukla H.D., Lasky S.R., Baliga N.S., Thorsson V., Sbrogna J.,
 RA Swartzell S., Weir D., Hall J., Dahl T.A., Welti R., Goo Y.A.,
 RA Leithauer B., Keller K., Cruz R., Danson M.J., Hough D.W.,
 RA Maddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,
 RA Isenbarger T.A., Peck R.F., Pohlschroder M., Spudich J.L., Jung K.-H.,
 RA Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,
 RA Ehardt H., Lowe T.M., Liang P., Riley M., Hood L., Dassarma S.;
 RT "Genome sequence of Halobacterium species NRC-1";
 RL Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).
 DR EMBL: AE005035; AAG19415.1; -;
 DR InterPro: IPR001395; -;
 DR Pfam: PF00248; aldo_ket_red; 1.
 SQ SEQUENCE 336 AA; 37376 MW; 04C7676CED9B8E04 CRC64;

Query Match 21.8%; Score 238; DB 1; Length 336;
 Best Local Similarity 33.5%; Pred. No. 5e-11;
 Matches 67; Conservative 34; Mismatches 81; Indels 18; Gaps 7;
 QY 1 DTPLEETMYALRDIVASGKALYVGISSYGPGLTAEAEFMAEERG-CPLLIHOPSYSIINR 57
 DB 141 DITARELMTLNDLVADGLVHVLGASTLQPNARARANELARAEWGEPFVLPQPRYNLV 200
 QY 58 NRWVEEPGDDGNNLQSAANGLGVIAFSPLAQGLLTKYL--DGIPGSRASQKSLSE 115
 DB 201 DREIE--GD-----YLTFAREQGLAVCPWSPLAQGLTGTGYSDRDLGTGESKASDSRWAD 254
 QY 116 GMLNVNNDIMVRKLNDAIQAERGQSLAQMALAWLREGEYGDVTSALIGASSVBOLEDN 175
 DB 255 TYLTDFANFADHVDLDAVEDEASPPQVALWLLHRDG-----ITAPIVGARTPDOLSE 308
 QY 176 SLDSLNNLEFSDAELEAIDE 195
 DB 309 NIDAA-TIDLTDDQMARLTD 327
 RESULT 10
 Q9PGD5 PRELIMINARY; PRT; 387 AA.
 AC Q9PGD5;
 DT 01-OCT-2000 (TremBLrel. 15, Created)
 DT 01-OCT-2000 (TremBLrel. 15, Last sequence update)

RESULT	11
ID	O59826
PRELIMINARY;	PRT; 344 AA.
AC	O59826;
DT	01-AUG-1998 (TREMBLrel. 07, Created)
DT	01-AUG-1998 (TREMBLrel. 07, Last sequence update)
DT	01-MAY-2000 (TREMBLrel. 13, Last annotation)

SQ SEQUENCE 354 AA; 38846 MW; 21D4074605E80EB6 CRC64;

BEST LOCAL SIMILARITY 32.3%; PRED. NO. 3.1E-09;
Matches 69; Conservative 35; Mismatches 66; Indels 42; Gaps 9;

```
QY 2 TPLEETMYALRDIVASGKALYGVGSSY-----GPELTAAEAFAEFMAEGCPLLIHQP 52
|||||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 134 TPLEETMEALDVLROGKVRIGCSNFTGWQIMKALGISEKDKRORFVSQ-----IH-- 186

QY 53 SYSIINRWEEPCDGENLLQSANNGLVIAFSPLAQGLLTDKYL--DGIPGSRASOG 110
|:  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 187 -YTL-----EADAIEYELLPISVDQGLVLIWSPLAGLLSGKHRNOSAPGSRQFAG 239

QY 111 KSL-----SEGMLNVNIDMKVRLNDIAQERGOSLAQMALAWLVRQEGYGADVTVSALI 165
|:  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 240 WTEPPYVDEERLWNI-----VDTLLSVADGRGVSAAQVALAWLI-----GRKAVTSIII 288

QY 166 GASSVQLDONSLSLNNLEFSDALEAIDEIS 197
|:  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 289 GGRTEAQFKDNL-AAADLQLSAEERKRLDDVS 319

RESULT 13
Q9HVV1 PRELIMINARY; PRT; 345 AA.
AC Q9HVV1;
DT 01-MAR-2001 (TReMBLrel. 16, Created)
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
DE 01-MAR-2001 (TReMBLrel. 16, Last annotation update)
DE PROBABLE OXIDOREDUCTASE.
GN PA4434.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PA01;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle D.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltry L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Latbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen.";
RL Nature 406:959-964(2000).
DR EMBL; AE004857; AAC07822.1; -.
DR InterPro; IPR001395; -.
DR Pfam; PF00248; aldo.ket.red; 1.
DR PRINTS; PR00069; ALDKETRDASE.
DR PRODom; PD002880; -. 1.
SQ SEQUENCE 345 AA; 38644 MW; 6866B69139E62508 CRC64;
```

```
Query Match 19.3%; Score 211; DB 2; Length 345;
Best Local Similarity 33.0%; Pred. No. 7e-09;
Matches 66; Conservative 33; Mismatches 85; Indels 16; Gaps 6;

QY 2 TPLEETMYALRDIVASGKALYGVGSSYGPPELTAAEAFAEFMAEGCPLLIHQPSYSIINRW 60
|||||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 154 TPLEETLEVLDEQVRAGKIRHIGLSNETPWGTFRLAEERGWPRAVSIQNPYLLNRS 213

QY 61 VPEPGDGENLLQSANNGLVIAFSPLAQGLLTDKYLDTGI-PEGSRASQGSLSGMLN 119
|:  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 214 FE-----VGLAEIATREQGLLAYSPMAFGMLSGRYADGARPANARISLYSRFTR-YTN 266

QY 120 VNNIDMKVRLNDIAQERGOSLAQMALAWLVRQEGYGADVTVSALIGASSVQLDONSLS 179
|:  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 267 POAEACARYVALAREHGLEPAQMALAYVT-----SRPFVTSNIIGATSLQLETNUGS 320

QY 180 LNNLEFSDALEAIDEISHD 199
|:  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 321 V-DLRDLEEVLAGIDAIHRE 339
```

```
RESULT 14
Q9F2Z5 PRELIMINARY; PRT; 330 AA.
AC Q9F2Z5;
DT 01-MAR-2001 (TReMBLrel. 16, Created)
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)
DE PUTATIVE OXIDOREDUCTASE.
GN SCD19.07.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Seeger K.J., Harris D.;
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Cerdeno A.M., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RX MEDLINE=97000351; PubMed=8843436;
RA Redenbach M., Kieser H.M., Denapalte D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL Mol. Microbiol. 21:77-96(1996).
DR EMBL; AL392149; CAC08308.1; -.
SQ SEQUENCE 330 AA; 36116 MW; 5733CFEF9F7BEEA5 CRC64;

Query Match 19.3%; Score 210.5; DB 2; Length 330;
Best Local Similarity 32.0%; Pred. No. 7.2e-09;
Matches 63; Conservative 36; Mismatches 81; Indels 17; Gaps 6;

QY 2 TPLEETMYALRDIVASGKALYGVGSSYGPPELTAAEAFAEFMAEG--CPLLIHQPSYSIINR 59
|||||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 136 TPEETIWAIDVLVQGGKILYAGSNFPQYKTAQANEIARRGGTIGLYSEQCLYNLAER 195

QY 60 WVEPEGDGENLLQSANNGLVIAFSPLAQGLLTDKYLDTGIPEGSRASQGSLSGMLN 119
|:  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 196 RAE-----MEVIPAARDYGLGVIPWSPHGLGGVKKKATQGRAS-GRA-ADALKD 247

QY 120 VNNIDMKVRLNDIAQERGOSLAQMALAWLVRQEGYGADVTVSALIGASSVQLDONSLS 179
|:  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 248 TKTREQIQAIEDLLDKHGLQPGEAALAWLLTRPG-----VTGPIVGPRTEQOLDALSRA 301

QY 180 LNNLEFSDALEAIDEI 196
|:  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 302 L-ELELSEELTSLDEI 317

RESULT 15
Q9L1S0 PRELIMINARY; PRT; 336 AA.
AC Q9L1S0;
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)
DE PUTATIVE OXIDOREDUCTASE.
GN SCE56.02.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
```


Search completed: September 13, 2001, 17:56:48
Job time: 298 sec

us-09-105-117i-2.rsp A₂(LysE)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 13, 2001, 17:52:15 ; Search time 15.26 Seconds
(without alignments)
484.875 Million cell updates

Title: US-09-105-117I-2

Perfect score: 1091
Sequence: 1 DTPLEETMYVALRDIVASGKA.....SHDAGINIWAKATDSKTREN 216

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

.....

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query		ID	Description
		Match	Length		
1	225	20.6	324	1 YAJQ_ECOLI	P77735 escherichia
2	193	17.7	326	1 YDJG_ECOLI	P77256 escherichia
3	182.5	16.7	310	1 YCKK_BACSU	P46905 bacillus su
4	180.5	16.5	346	1 TAS_ECOLI	P46933 escherichia
5	178	16.3	340	1 YAKC_SCHPO	Q09923 schizosacch
6	161.5	14.8	323	1 YM98_MYCTU	Q50668 mycobacteri
7	159.5	14.6	330	1 AR72_HUMAN	O43488 homo sapien
8	157.5	14.4	331	1 AR73_HUMAN	O43488 homo sapien
9	156.5	14.3	327	1 AR71_RAT	O95154 homo sapien
10	152.5	14.0	387	1 STCV_EMENI	P38918 rattus norv
11	150	13.7	307	1 A115_TOBAC	Q00727 emericella
12	146	13.4	306	1 YQKF_BACSU	P04691 nicotiana t
13	146	13.4	310	1 IOLS_BACSU	P54569 bacillus su
14	145.5	13.3	331	1 GS69_BACSU	P46336 bacillus su
15	139.5	12.8	388	1 NORA_ASPLP	P80874 bacillus su
16	138.5	12.7	388	1 NORA_ASPPA	Q00049 aspergillus
17	133	12.2	285	1 ALKE_BABBO	Q00258 aspergillus
18	130.5	12.0	288	1 AADA_YEAST	P40690 babesia bov
19	127.5	11.7	329	1 AADA_YEAST	P47182 saccharomyc
20	122.5	11.2	376	1 AADE_YEAST	Q07747 saccharomyc
21	119.5	11.0	286	1 YDBC_ECOLI	P42884 saccharomyc
22	119.5	11.0	385	1 AAD_PHACH	P25906 escherichia
23	108	9.9	282	1 Y166_YEAST	Q01752 phanerochae
24	107.5	9.9	310	1 SPFD_MALDO	P47137 saccharomyc
25	106	9.7	318	1 XYLL_PICST	P28475 malus domes
26	106	9.7	363	1 AAD3_YEAST	P31867 picchia stip
27	103.5	9.5	306	1 IN22_MAIZE	P25612 saccharomyc
28	96.5	8.8	295	1 MORA_PSEPU	P49249 zea mays (m
29	95.5	8.8	271	1 YB1I_ECOLI	Q02198 pseudomonas
30	91.5	8.4	381	1 DCOP_TRIRE	P75809 escherichia
31	91	8.3	835	1 CD97_HUMAN	P21594 trichoderma
32	90.5	8.3	1252	1 PG22_HUMAN	P48960 homo sapien
33	89.5	8.2	379	1 DCOP_TRITHA	P16885 homo sapien
34	89.5	8.2	379	1 DCOP_TRITHA	Q12709 trichoderma

ALIGNMENTS

RESULT	1	
YAJO_ECOLI		
ID	YAJO_ECOLI	STANDARD; PRT; 324 AA.
AC	P77735;	
DT	01-NOV-1997 (Rel. 35, Created)	
DT	15-DEC-1998 (Rel. 37, Last sequence update)	
DT	15-DEC-1998 (Rel. 37, Last annotation update)	
DE	HYPOTHETICAL OXIDOREDUCTASE IN PGPA-ISPA INTERGENIC REGION.	
GN	YAJO.	
OS	Escherichia coli.	
OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;	
OC	Escherichia.	
OX	NCBI_TaxID=562;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=K12 / MG1655;	
RX	MEDLINE=97426617; PubMed=9278503;	
RA	Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,	
RA	Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,	
RA	Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,	
RA	Mau B., Shao Y.;	
RT	"The complete genome sequence of Escherichia coli K-12.";	
RL	Science 277:1453-1474(1997).	

```

[2]
RN      ROBERTS D., ALLEN E., ARAUJO R., APARICIO A., CHUNG E., DAVIS K.,
RP      DUNCAN M., FEDERPIEL N., HYMAN R., KALMAN S., KOMP C., KURDI O.,
RA      LEW H., LIN D., NAMATH A., OEFNER P., SCHRAMM S., DAVIS R.W.;
RR      Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
RRL     ! - SIMILARITY: BELONGS TO THE ALDO/KETO REDUCTASE 2 FAMILY.
RC
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CCC     between the Swiss Institute of Bioinformatics and the EMBL outstation
CCC     the European Bioinformatics Institute. There are no restrictions on its
CCC     use by non-profit institutions as long as its content is in no way
CCC     modified and this statement is not removed. Usage by and for commercial
CCC     entities requires a license agreement (See http://www.isb-sib.ch/announce/
CCC     or send an email to license@isb-sib.ch).
CC
EMBL: AE000148; AAC73522.1; ALT INIT.
DR      EMBL: U82664; AABA0175.1; ALT_INIT.
DR      EcoGene, Egi3611; yajO.
DR      Hypothetical protein; Oxidoreductase.
KW      ACT_SITE 127 127 HYDROGEN-BOND DONOR (PROBABLE).
RQ      SEQUENCE 324 AA; 36420 MW; 0B69E09F2BEDF9B1 CRC64;

```

```
Query Match      20.6%; Score 225; DB 1; Length 324;
Best Local Similarity 31.9%; Pred.No. 4.2e-11;
Matches 60; Conservative 39; Mismatches 63; Indels 26; Gaps 6;

QOY 1 DTPLEETMYALRDIVASGRKALVYGISSYPELTAEAAEFMAEEGCPCLLIH-QSYSLINR 59
      :|||::||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::
DDB 132 NPTIELTALNDVVKGARKARIGASSHASAQFALELQKHGWQAQFVSMODHYNLIIYR 191

QOY 60 WVEEPDDCENILQSAAANGVLGVIAESPLAQQGLTLTDKYLDIGIPEGRASQ - - - - GKSL 113
```

```

Db 192 -----EEREMPLCYQGVAVIPWPLARGLTR-----PWGETTARLVSDVGVKNL 239
QY 114 SEGMLNNDVNRKLNDAQERQSLAQMALAWLVRQEGYADTVTSALIGASSVEQL 173
Db 240 YKES-DENDAQAERTLTGVSEELGATRAQVALWLLSKFG-----IAPIITGTSREQL 292
QY 174 DNSLDSLN 181
Db 293 DELLNAVD 300

RESULT 2
YDJG_ECOLI
ID YDJG_ECOLI STANDARD; PRT; 326 AA.
AC P77256;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE HYPOTHETICAL OXIDOREDUCTASE IN ANSA-RND INTERGENIC REGION.
GN YDJG.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RA Alba H., Baba T., Fujita K., Hayashi K., Honjo A., Horiuchi T.,
RA Ikemoto K., Inada T., Isono K., Itoh T., Kanai K., Kasai H.,
RA Kashimoto K., Kim S., Kimura S., Kitagawa M., Kitakawa M., Makino K.,
RA Masuda S., Miki T., Mizobuchi K., Mori H., Motomura K., Nakamura Y.,
RA Nishimoto H., Nishio Y., Oshima T., Saito N., Sampel G., Seki Y.,
RA Tagami H., Takemoto K., Wada C., Yamamoto Y., Yano M.;
RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
CC !- SIMILARITY: BELONGS TO THE ALDO/KETO REDUCTASE 2 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE000272; AAC74841.1; -.
CC EMBL; D90821; BAA15569.1; -.
CC EMBL; D90820; BAA15562.1; -.
CC EcoGene: EG13483; ydjG.
KW Hypothetical protein; Oxidoreductase.
FT ACT_SITE 136 136 HYDROGEN-BOND DONOR (PROBABLE).
SQ SEQUENCE 326 AA; 36328 MW; 1FF5C5252708978 CRC64;

```

```

Query Match 17.7%; Score 193; DB 1; Length 326;
Best Local Similarity 31.0%; Pred. No. 1.6e-08;
Matches 62; Conservative 40; Mismatches 72; Indels 26; Gaps 7;

QY 2 TPLEETWALRDIVASGKALYVGISSYGPELTAAAEAFMAEEGCPLLIHQPSYIINRWV 61
Db 145 TPIAETVAVLNKLSEKGRKIRAGANVDADHIREVLYQY-----GELDIQAKYSILDRAM 199
QY 62 EEPGDDGNLLOSAANGLGVAFSPLAQGLTLDKYL-DGIEGSRASOGKSLSEGMLNV 120

```

```

Db 200 E-----NELLPLCRDNGIWWQVSPLEQGLLTGTITRDVPGGRANKVWFQRENMLKV 253
QY 121 NNIDVNRKLNDAQERQSLAQMALAWLVRQEGYADTVTSALIGASSVEQLDNLDSL 180
Db 254 --IDMLEQWQPLCARYQCTIPTLALAWILKQ-----SDLISLGATAPAEQVRENVAAL 305
QY 181 N-NLEFSDAEL-----EAJD 194
Db 306 NINLSDADATLREMAEAL 325

RESULT 3
YCKC_BACSU
ID YCKC_BACSU STANDARD; PRT; 310 AA.
AC P46905;
DT 01-NOV-1995 (Rel. 32, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE HYPOTHETICAL OXIDOREDUCTASE IN NATB-RAPJ INTERGENIC REGION (ORFC).
GN YCKC.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RA Kumano M., Tanakoshi A., Yamane K.;
RT "A 32 kb nucleotide sequence from the region of the lincomycin-
RT resistance gene (22-25 degree) of the Bacillus subtilis chromosome and
RT identification of the site of the lin-2 mutation.";
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 1-268 FROM N.A.
RC STRAIN=BD99;
RA Cheng J., Guffanti A.A., Krulwich T.A.;
RL Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.
CC !- SIMILARITY: BELONGS TO THE ALDO/KETO REDUCTASE 2 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AB000617; BAA22238.1; -.
CC EMBL; Z99105; CAB12071.1; -.
CC EMBL; U30873; AAB53024.1; -.
CC HSP: P80276; IAH3.
CC SubtilList; BG11340; yckC.
CC InterPro; IPR001395; -.
CC Pfam; PF00248; aldo_ket_red; 1.
KW Hypothetical protein; Oxidoreductase.
FT ACT_SITE 127 127 HYDROGEN-BOND DONOR (PROBABLE).
SQ SEQUENCE 310 AA; 34800 MW; C06BF4195D25C91C CRC64;

```

```

Query Match 16.7%; Score 182.5; DB 1; Length 310;
Best Local Similarity 29.6%; Pred. No. 1e-07;
Matches 58; Conservative 38; Mismatches 79; Indels 21; Gaps 7;

```

```

QY 2 TPLEETWALRDIVASGKALYVGISSYGPELTAAAEAFMAEEGCPLLIHQPSYIINRWV 61
Db 133 TPLAEVAGTLLKELDKGKIKAGASNLQYQ---QLODFNADG--YLEVFQAEYSLIQR-- 185
QY 62 EEPGDDGNLLOSAANGLGVAFSPLAQGLTLDKYL-DGIEGSRASOGKSLSEGMLNV 120
Db 186 ----DAERELLYCEKQGISFIPYPLASGLTGKTFQDTVFDDFR--KDKQFQGETFI 239
QY 121 NNIDVNRKLNDAQERQSLAQMALAWLVRQEGYADTVTSALIGASSVEQLDNLDSL 180

```

Db 240 HNLKVKDKLRAVAEEKQADTAHVALWLLTR-----PAIDAIIPGAKRPQLODNKLT 293

QY 181 NNLFEFSDAELEFAIDEI 196

Db 294 -NIELTEDEVNFISDI 308

RESULT 4

TAS_ECOLI ID TAS_ECOLI STANDARD; PRT; 346 AA.

AC Q46933; 1998 (Rel. 36, Created)

DT 15-JUL-1998 (Rel. 36, Last sequence update)

DT 15-DEC-1998 (Rel. 37, Last annotation update)

DE TAS PROTEIN.

GN TAS.

OS Escherichia coli.

OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

OC Escherichia.

OX NCBI_TaxID=562;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-K12 / MG1655;

RX MEDLINE=97426617; PubMed=9278503;

RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,

RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,

RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,

RA Mau B., Shao Y.;

RT "The complete genome sequence of Escherichia coli K-12.";

RL Science 277:1453-1474 (1997).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=B/R / WU3610;

RX MEDLINE=98240727; PubMed=9560382;

RA Timms A.R., Bridges B.A.;

RT "Reversion of the tyrosine ochre strain Escherichia coli WU3610 under

RT starvation conditions depends on a new gene tas.";

RL Genetics 148:1627-1635 (1998).

CC -1- SIMILARITY: BELONGS TO THE ALDO/KETO REDUCTASE 2 FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (see http://www.isb-sib.ch/announce/

CC or send an email to license@isb-sib.ch).

CC -----

CC EMBL; U29581; AAB40481.1; -

DR EMBL; AE000367; AAC75873.1; -

DR EMBL; Y14609; CAA74961.1; -

DR HSP; P14550; 2ALR

DR EcGene; EG13093; tas.

DR InterPro; IPR001395; -

DR Pfam; PF00248; aldo_ket_red; 1.

KW Oxidoreductase.

FT ACT_SITE 132 132 HYDROGEN-BOND DONOR (BY SIMILARITY).

SQ SEQUENCE 346 AA; 38499 MW; 111692D06CA07CD7 CRC64;

Query Match 16.5%; Score 180.5; DB 1; Length 346;

Best Local Similarity 30.1%; Pred. No. 1.7e-07;

Matches 63; Conservative 42; Mismatches 63; Indels 41; Gaps 10;

QY 4 LEETMYALRDIVASGKALYGVIS----SYGPDLTAEEAFMAEEGCPLLIH-QPSYSIINR 59

Db 157 LLDTDLALAEYQKAKIRYIGVSNFETAGVMRYLHLAD---KHLPRIVTQNPYSILNR 213

QY 60 WVEPGDDGNNLQSAANGLGVTAFAFSLAQLGLTDKYLGI-PEGSRAS-----QG 110

Db 214 SFE-----VGLAEVSQYEGVELLAYSCLGTLGTLGNKAPGARGNTLFSRFTYSG 267

QY 111 KSLSEGMLNNVNDIMVRKLNLDIAQERGQSLAQMALAWLVRQEGYGADTVTSALIGASSV 170

Db 268 EOTCKA-----VAAYVDIARRHGLDPAQMALAFVRRQ-----PFVASTLLGATTM 312

QY 171 EQDNLSDLSLNNLEFSD---AELEAIDEI 196

Db 313 DQKLTNIESL-HLELSEDLVLAIEIAVHQV 340

RESULT 5

YAKC_SCHPO ID YAKC_SCHPO STANDARD; PRT; 340 AA.

AC Q09923;

DT 01-FEB-1996 (Rel. 33, Created)

DT 01-FEB-1996 (Rel. 33, Last sequence update)

DT 01-FEB-1996 (Rel. 33, Last annotation update)

DE HYPOTHETICAL 37.7 KDA PROTEIN C1F7.12 IN CHROMOSOME I.

GN SPAC1F7.12 OR SPAC21E11.01.

OS Schizosaccharomyces pombe (Fission yeast).

OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;

OC Schizosaccharomycetales; Schizosaccharomycetaceae;

OC Schizosaccharomycetes.

OX NCBI_TaxID=4896;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=972;

RA Murphy L., Niblett D., Harris D., Barrell B.G., Rajandream M.A.,

RA Walsh S.V.;

RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.

CC -1- SIMILARITY: BELONGS TO THE ALDO/KETO REDUCTASE 2 FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (see http://www.isb-sib.ch/announce/

CC or send an email to license@isb-sib.ch).

CC -----

CC EMBL; Z67998; CAA91959.1; -

DR EMBL; Z67999; CAA91961.1; -

DR HSP; P14550; 2ALR

DR InterPro; IPR001395; -

DR Pfam; PF00248; aldo_ket_red; 1.

KW Hypothetical protein; Oxidoreductase.

FT ACT_SITE 126 126 HYDROGEN-BOND DONOR (PROBABLE).

SQ SEQUENCE 340 AA; 37710 MW; A4A9686A70968F06 CRC64;

Query Match 16.3%; Score 178; DB 1; Length 340;

Best Local Similarity 26.8%; Pred. No. 2.7e-07;

Matches 55; Conservative 42; Mismatches 88; Indels 20; Gaps 8;

QY 1 DTPLEETMYALRDIVASGKALYGVIS--SYGPDLTAEEAFMAEEGCPLLIHQPSYSIINRW 60

Db 131 ETPIEKIMGALKCKVEAGKIRYIGLSECSANTIRRAAVY-----PVSAYQVEYSPFSL 185

QY 61 VEEPGDDGNNLQSAANGLGVTAFAFSLAQLGLTDKYL--DGIPEG-SASQSGKSLSEGM 117

Db 186 IERP---EIGVMKACRENNITIVCYAPLGRGLTGAKYKSPDPEGDFERRKAPYQENF 242

QY 118 LNVNNDIMVRKLNLDIAQERGQSLAQMALAWLVRQEGYGADTVTSALIGASSVBQLDNSL 177

Db 243 --YKMLELVTKIEKIATANNITPGQLSLAWLLAQ---GDDIL--PIPGTKRVKYLENF 294

QY 178 DSLNNLEFSDAELEAIDEISHDAGI 202

Db 295 GAL-KVKLSDATVKEIREACDRAEV 318

RESULT 6

YM98_MYCTU ID YM98_MYCTU STANDARD; PRT; 323 AA.

```

GN AKR7A2 OR AFAR OR AKR7.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=98244807; PubMed=9576847;
RA Ireland L.S., Harrison D.J., Neal G.E., Hayes J.D.;
RT "Molecular cloning, expression and catalytic activity of a human AKR7
RT member of the aldo-keto reductase superfamily: evidence that the
RT major 2-carboxybenzaldehyde reductase from human liver is a homologue
RT of rat aflatoxin B1-aldehyde reductase.";
RL Biochem. J. 332:21-34(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=99040634; PubMed=9823300;
RA Praml C., Savelyeva L., Perri P., Schwab M.;
RT "Cloning of the human aflatoxin B1-aldehyde reductase gene at lp35-
RT p36.1 in a region frequently altered in human tumor cells.";
RL Cancer Res. 58:5014-5018(1998).
RN [3]
RP SEQUENCE FROM N.A.
RC Hall R.;
RA Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
CC -!- FUNCTION: CAN METABOLIZE THE AFLATOXIN B1 (AFB1) PROTEIN-BINDING
CC DIHYDRODIOL BY FORMING NONBINDING AFB1 DIALCOHOL. COULD BE
CC INVOLVED IN PROTECTION OF LIVER AGAINST THE TOXIC AND CARCINOGENIC
CC EFFECTS OF AFLATOXIN B1. ACTS AS A 2-CARBOXYBENZALDEHYDE
CC REDUCTASE.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -!- SIMILARITY: BELONGS TO THE ALDO/KETO REDUCTASE 2 FAMILY.
CC -----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
EMBL; AF026947; AAC52104.1; -.
DR DR EMBL; Y16675; CAAT6347.1; -.
DR DR EMBL; AL035413; CAB72321.1; -.
DR DR MIM; 603418; -.
KW Oxoaldereductase.
FT ACT_SITE 112 112 HYDROGEN-BOND DONOR (PROBABLE).
FT CONFLICT 113 113 A -> T (IN REF. 1).
SQ SEQUENCE 330 AA; 36618 MW; 3BFEB7ED0CAFAD54 CRC64;
-----
Query Match 14.6%; Score 159.5; DB 1; Length 330;
Best Local Similarity 27.8%; Pred. No. 8e-06;
Matches 59; Conservative 33; Mismatches 95; Indels 25; Gaps 7;

QY 2 TPLEETMYALRDIVASGKALYVCISGYSGELTAEEAEFFMAEECPL-LHQPSYSINRW 60
DB ||:|||::: || :||:|: | | :|| ::| |
118 TPVEETHLCQRHLQHKGKFVELGSLNSWAEVAEICTLCKSNSGWILLPTVYGYNATTRQ 177
QY || VPEPGDDGNNLLQSAANGLGVIAFSPLAGOLLTKYLDTGIPIGRASQ---GKSLSSEG 117
DB || :|| :|| ||||| ||||| ||||| :|| :||
178 VE-----TELFPCLRHFGFLRVFYNPFLAGLLTGKYKKYEDKDGGKPVGRRFGNSWAETY 231
QY 118 LN-----VNNDIMVKLNIDTAO-ERGSOLAQMALAWLRE----OGEVGADVTSALIG 166
DB || :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
232 RNRFWRKEHHFEATALYEKAQLQAAYGASAPSVTASLRWMYVHHSLOQGAHG---DAVLG 287
QY 167 ASSVEOLDNLSLDNLNEFSDAELEIDEISH 198
DB ||:|||::: | :|| :|| :|| :|| :||
288 MSSLEGLEONLAATEEGPLEPAVVADFNAQWH 319
```

```
RESULT 8
AR73_HUMAN STANDARD; PRT; 331 AA.
AC O95154; O9NUC3;
DT 01-OCT-2000 (Rel. 40, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE AFLATOXIN B1 ALDEHYDE REDUCTASE 2 (EC 1.-.-.-) (AFB1-AR 2).
GN AKR7A3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID=9606;
RN SEQUENCE FROM N.A., AND CHARACTERIZATION.
RP TISSUE=Liver;
RX MEDLINE=93315412; PubMed=10383892;
RA Knight L.P., Primiano T., Groopman J.D., Kensler T.W., Sutter T.R.;
RT "cDNA cloning, expression and activity of a second human aflatoxin
RT B1-metabolizing member of the aldo-keto reductase superfamily,
RT AKR7A3."
RL Carcinogenesis 20:1215-1223(1999).
RN SEQUENCE FROM N.A.
RP Hall R.;
RA Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: CAN METABOLIZE THE AFLATOXIN B1 (AFB1) PROTEIN-BINDING
CC DIHYDRODIOL BY FORMING NONBINDING AFB1 DIALCOHOL. COULD BE
CC INVOLVED IN PROTECTION OF LIVER AGAINST THE TOXIC AND CARCINOGENIC
CC EFFECTS OF AFLATOXIN B1.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -1- SIMILARITY: BELONGS TO THE ALDO/KETO REDUCTASE 2 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF040639; AAD02195.1; -
CC EMBL; AL035413; CAB72322.1; -
CC InterPro; IPR001395; -
CC PRINTS; PR00069; ALDKETREDASE.
CC OXidoreductase.
CC ACT SITE 113 113 HYDROGEN-BOND DONOR (PROBABLE).
CC CONFLICT 51 51 E -> D (IN REF. 1).
CC CONFLICT 138 138 V -> M (IN REF. 1).
CC CONFLICT 201 201 A -> ADQSPGCGSFGWTGLPGADCCFPS (IN REF.
CC 2).
CC CONFLICT 215 215 N -> D (IN REF. 1).
CC CONFLICT 323 323 T -> A (IN REF. 1).
CC SEQUENCE 331 AA; 37206 MW; B9C32C33C7102AB3 CRC64;

Query Match 14.4%; Score 157.5; DB 1; Length 331;
Best Local Similarity 27.8%; pred. No. 1.2e-05;
Matches 59; Conservative 30; Mismatches 98; Indels 25; Gaps 7;

QY 2 TPLEETMYALRDIVASKALYVGISYSGPELTAEAEFMAEEGCL-LIHQPSYSINRW 60
|||:| | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 119 TPVETLRACHQLHQEGKVELGSLNSAAWEAEICTLCKSNGLTPTVYOGMYNAITRQ 178
|||:| | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 61 VEEPDGGENLQSAANGLGVIAFSLAQGLLTDKLYDGIPEGSRAQ----GKSLSEG 117
|| | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 179 VE-----TELPFLCRHFLGRFYAFNPLAGLLTGKTKYEDKNGKQVPVGRFGNTWAEY 232
|||:| | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 118 LN-----VNNIDMVRKLNIDIAQ-ERQSLAQMALAWLRLRE---QGEYGADTVTSALIG 166
|| | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 233 RNRWKKEHHFEGIALVEKALQAYGASAPSWTSATLRWYTHHSQLOGAHG-----DAVLIG 288
|| | | | | | | | | | | | | | | | | | | | | | | | | | |
```

ID	YQKF_BACSU	STANDARD;	PRT;	306 AA.
AC	P54569;			

DT 01-OCT-1996 (Rel. 34, Created)


```

DT 01-OCT-1996 (Rel. 34, Last sequence update)
DE 15-JUL-1998 (Rel. 36, Last annotation update)
DF HYPOTHETICAL OXIDOREDUCTASE IN ANSR-BMRU INTERGENIC REGION.
GN YQKF.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=168 / JH642;
RA Kobayashi Y., Mizuno M., Masuda S., Takemaru K., Hosono S.,
RA Sato T., Takeuchi M.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE ALDO/KETO REDUCTASE 2 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; D84432; BAA12638.1; -
DR EMBL; Z99116; CAB14294.1; -
DR Subtilist; BG11761; YQKF.
DR InterPro; IPR001395; -
DR Pfam; PF00248; aldo_ket_red; 1.
KW Hypothetical protein; Oxidoreductase.
FT ACT_SITE 119 119 HYDROGEN-BOND DONOR (PROBABLE).
SQ SEQUENCE 306 AA; 34717 MW; CD79B13B956B000F CRC64;

Query Match 13.4%; Score 146; DB 1; Length 306;
Best Local Similarity 25.58; Pred. No. 8.9e-05;
Matches 50; Conservative 44; Mismatches 74; Indels 28; Gaps 8;

QY 4 LEETMYALRDIVASGKALYGVISYGPETAFAAEFMAEGCPLLIHQPSYSIINRWEE 63
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
127 :DETEAFEEELKQNGVIRYGISSIRPNV---IKYVKKSIVSIMQ--PSLFDRRPEE 181
QY 64 PGDDGENLLQSAANGLGVIAFSPAQLGLTDKYLDGIPGSRASQKSLSEGLMNVNN 122
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
182 -----WLPLEBHQISVARGVAVKGLTEKPLDQASESMKONGVLSYFEELTNAR- 233
QY 123 IDMYRKLNDIAQERGQSLAQMALAWLREOGEYCADVTSTALIGASSVEQLDNLN 182
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
234 ----KAMEEVAPD--LSMTEKSLQYLIAQ-----PAVASVITGASKIEQLRNIQAANA 281
QY 183 LEFSDAELEAIDEISH 198
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
282 RRLTEEEIKALQ--SH 295

RESULT 13
IDLS_BACSU STANDARD; PRT; 310 AA.
AC P46336;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE IOLS PROTEIN (VEGETATIVE PROTEIN 147) (VEG147).
GN IOLS OR SS92ER.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=168 / BGSC1A1;
RX MEDLINE=96093926; PubMed=7584049;
Yoshida K.-I., Seki S., Fujimura M., Miwa Y., Fujita Y.;

RT "Cloning and sequencing of a 36-kb region of the Bacillus subtilis
RT genome between the gnt and iol operons.";
RL DNA Res. 2:61-69(1995).
RN [2]
RP SEQUENCE FROM N.A.
RA STRAIN=168 / BGSC1A1;
RA Fujita Y., Shibayama T., Ishio I., Aoyama D., Yoshida K.-I.;
RT "Organization and transcription of the myo-inositol operon, iol, of
RT Bacillus subtilis.";
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 1-20.
RA STRAIN=IS58;
RX MEDLINE=97443988; PubMed=9298659;
RA Antelmann H., Bernhardt J., Schmid R., Mach H., Voelker U.,
RA Hecker M.;
RT "First steps from a two-dimensional protein index towards a response-
RT regulation map for Bacillus subtilis.";
RL Electrophoresis 18:1451-1463(1997).
CC -1- PATHWAY: MYO-INOSITOL CATABOLISM.
CC -1- SIMILARITY: BELONGS TO THE ALDO/KETO REDUCTASE 2 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AB005554; BAA21607.1; -
DR EMBL; Z99124; CAB16014.1; -
DR HSSP; P23457; ILWI.
DR Subtilist; BG11363; iols.
DR Pfam; PF00248; aldo_ket_red; 1.
KW Oxidoreductase.
FT ACT_SITE 125 125 HYDROGEN-BOND DONOR (PROBABLE).
SQ SEQUENCE 310 AA; 35168 MW; A870F26F8684867 CRC64;

Query Match 13.4%; Score 146; DB 1; Length 310;
Best Local Similarity 25.18; Pred. No. 9e-05;
Matches 49; Conservative 37; Mismatches 83; Indels 26; Gaps 7;

QY 2 TPLEETMYALRDIVASGKALYGVISYGPETAFAAEFMAEGCPLLIHQPSYSIINRW 61
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
131 TPKDEAVNALNEMKKAKIRSIGVSNFSLQKKA---NKDGL-VDVLQGEYNLLNREA 185
QY 62 EEPGDDGENLLQSAANGLGVIAFSPAQLGLTDKYLDG--IPGSRASQKSLSEGLMLN 119
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
186 E-----KTFPPYTKHEINISFIPYPLVSLGGLAGKYTEDTTFPEGLRNEQEHF-KGERF 238
QY 120 VNNIDMYRKLNDIAQERGQSLAQMALAWLREOGEYCADVTSTALIGASSVEQLDNLN 179
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
239 KENIRKYNKLAIPAEKHNVDIPHIVLAWYL-----ARPEIDILIPGAKRADQL----- 286
QY 180 LNNLEFSDAELEAID 194
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
287 IDNKTADVTLSQED 301

RESULT 14
IDGS69_BACSU STANDARD; PRT; 331 AA.
AC P80874; Q07583;
DT 01-NOV-1997 (Rel. 35, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE GENERAL STRESS PROTEIN 69 (GSP69).
GN YHDN.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.

```

OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RA Noback M.A., Terpstra P., Holsappel S., Venema G., Bron S.;
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 1-25.
RC STRAIN=IS58;
RX MEDLINE=97443988; PubMed=9298659;
RA Antelmann H., Bernhardt J., Schmid R., Mach H., Voelker U.,
RA Hecker M.;
RT "First steps from a two-dimensional protein index towards a response-
regulation map for Bacillus subtilis";
RL Electrophoresis 18:1451-1463(1997).
CC -I- INDUCTION: BY HEAT SHOCK, SALT STRESS, OXIDATIVE STRESS, GLUCOSE
LIMITATION AND OXYGEN LIMITATION.
CC -I- SIMILARITY: BELONGS TO THE ALDO/KETO REDUCTASE 2 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
or send an email to license@sib-sib.ch).
CC
CC EMBL; Y14082; CAA74498.1; -
DR EMBL; 299109; CAB12792.1; -
DR HSSP; P23457; ILWI.
DR Subtilisin; BG13020; yhdN.
DR Pfam; PF00248; aldo_ket_red; 1.
KW Oxidoreductase; Heat shock.
FT ACT_SITE 125 125 HYDROGEN-BOND DONOR (PROBABLE).
FT CONFLICT 25 25 G -> K (IN REF. 2).
SQ SEQUENCE 331 AA; 37312 MW; 82BC24D46E4994D0 CRC64;

Query Match 13.3%; Score 145.5; DB 1; Length 331;
Best Local Similarity 26.1%; Pred. No. 0.00011;
Matches 57; Conservative 37; Mismatches 91; Indels 33; Gaps 8;

QY 3 PLEETMYALRDIVASGKALVYGVISGPELTAEAAEFMAEEGCPLLIHQPSYSIINRW 62
Db 132 PLEETMYALRDIVASGKALVYGVISGPELTAEAAEFMAEEGCPLLIHQPSYSIINRW 62
QY 63 EPDGDGNIQSAANGLGVIAFSPLAQGLLTOKYL-----DGIPEGSRASOGKSLSE 115
Db 187 -----ESVLPYAKDKNITLLYGLRGLLTCKMTTEYTFEGDDLRNHPKQKPRFE 240
QY 116 GMLNVNMDVRKLNDAERGOSLAQMALAWLVREOGEYGDVTVSALIGASSVQDLN 175
Db 241 YLSAVNQLDKLAK-----TRYGKSVIHLAVRWILDQP---GADI---ALWGARKPGQLEA 289
QY 176 SLD-----SLNNLEFSDAELEAIDEISHDAGINIWAKAT 209
Db 290 LSEITGWTNLNSEDQKDINTILENTISDPVGPPEFMAPPT 327

RESULT 15

NORA ASPFL
ID NORA ASPFL STANDARD; PRT; 388 AA.
AC Q00049;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE NORSOLORINIC ACID REDUCTASE (EC 1.1.1.-).
GN NORA OR ADH-2.
OS Aspergillus flavus.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
OX NCBI_TaxID=5059;
RN [1]

RP SEQUENCE FROM N.A.
RC STRAIN=NRRL 3357;
RX MEDLINE=96156784; PubMed=8593042;
RA Cary J.W., Wright M., Bhatnagar D., Lee R., Chu F.;
RT "Molecular characterization of an Aspergillus parasiticus
dehydrogenase gene, nora, located on the aflatoxin biosynthesis gene
cluster";
RL Appl. Environ. Microbiol. 62:360-366(1996).
CC -I- PATHWAY: AFLATOXIN BIOSYNTHESIS.
CC -I- SIMILARITY: BELONGS TO THE ALDO/KETO REDUCTASE 2 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
or send an email to license@sib-sib.ch).
CC
CC EMBL; U32377; AAC49167.1; -
DR EMBL; 148 148 HYDROGEN-BOND DONOR (POTENTIAL).
KW Oxidoreductase.
FT ACT_SITE 388 AA; 43751 MW; 758B8187187BF5B0 CRC64;
SQ SEQUENCE 388 AA; 43751 MW; 758B8187187BF5B0 CRC64;

Query Match 12.8%; Score 139.5; DB 1; Length 388;
Best Local Similarity 25.1%; Pred. No. 0.00041;
Matches 49; Conservative 39; Mismatches 84; Indels 23; Gaps 7;

QY 2 PLEETMYALRDIVASGKALVYGVISGPELTAEAAEFMAEEGCPLLIHQPSYSIINRW 60
Db 154 TSVEEYMRSLNHLVANGKVLGLVSDTPAWLVVKNFAFANGLTFFSVYQGHWSAFA 212
QY 61 VEEPGDGNILQSAANGLGVIAFSPLAQGLLTOKYL-----DGIPEGSRASOGKSLSEGLN 120
Db 213 -----DPERDILPMCESEGWGLAPWGLVGRG-----QFRSAEEFSR--EGRKM--GPQDE 258
QY 121 NNIDMYRKLNDIAERGOSLAQMALAWLVREOGEYGDVTVSALIGASSVQDLN 180
Db 259 KHRRLEKLDQMAQKQKTRKATSIQAQVVMHKA-----PYVFPVIGRKVEHLKENIEAL 312
QY 181 NNLEFSDAELEAIDE 195
Db 313 -GLVLEEEIREIDD 326

Search completed: September 13, 2001, 17:57:09
Job time: 294 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 13, 2001, 17:34:55 ; Search time 25.14 Seconds
(without alignments)
654.483 Million cell updates/sec

Title: US-09-105-117I-2
 Perfect score: 1091
 Sequence: 1 DTPLEETMYALRDIVASGKA.....SHDAGINTWAKATDSKTREN 216

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

```

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                  Maximum Match 10%
                  Listing first 45

```

```
Database :      PIR_68:*
1:  pir1:*
2:  pir2:*
3:  pir3:*
4:  pir4:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query			DB	ID	Description
	Score	Match	Length			
1	529.5	48.5	346	2	G65086	hypothetical prote
2	526.5	48.3	346	2	E85959	probable reductas
3	449	41.2	319	2	T33337	probable ion chann
4	366	33.5	287	2	B86873	oxidoreductase yug
5	283.5	26.0	315	2	A75289	probable potassium
6	267	24.5	319	2	H72391	K+ channel, beta s
7	257	23.6	336	2	T44988	oxidoreductase [im
8	238	21.8	336	2	K84256	probable oxidoredu
9	237.5	21.8	387	2	F82815	voltage-gated pota
10	227.5	20.9	344	2	T41659	probable potassium
11	225	20.6	324	2	E85538	probable NAD(P)H-d
12	225	20.6	348	2	C64771	probable oxidoredu
13	213	19.5	354	2	T44934	mocA protein [impo
14	211	19.3	345	2	B83093	probable oxidoredu
15	209	19.2	330	2	T07394	probable potassium
16	209	19.2	367	2	A53131	alpha-dendrotoxin-
17	208	19.1	367	2	S66502	potassium channel
18	208	19.1	367	2	A53112	RCKbeta2 protein -
19	206.5	18.9	316	2	F83170	probable oxidoredu
20	205	18.8	401	2	S66503	potassium channel
21	205	18.8	408	2	I59393	potassium channel
22	205	18.8	419	2	I55463	K+ channel beta-su
23	197.5	18.1	326	2	D85787	hypothetical prote
24	197	18.1	352	2	C82294	oxidoreductase Tas
25	196	18.0	328	2	D84315	aryl-alcohol dehyd
26	195	17.9	324	2	T52133	potassium channel
27	194.5	17.8	316	2	F69978	sugar-phosphate de
28	193	17.7	326	2	C64937	hypothetical prote
29	193	17.7	327	2	E75296	aldo/keto reductas

ALIGNMENTS

RESULT	1
--------	---

G65086
hypothetical protein b3001 - Escherichia coli (strain K-12)
C:Species: Escherichia coli
C:date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 18-Feb-2000
C:Accession: G65086
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;
A.: Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617
A:Accession: G65086
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:molecule type: DNA
A:Residues: 1-346 <BLAST>
A:Cross-references: GB:AE000382; GB:U00096; NID:g2367182; PIDN:AAC76037.1; PID:gl7893
A:Experimental source: strain K-12, substrain MG1655
C:Superfamily: fission yeast pyridoxine 4-dehydrogenase

Query Match

Best Local Similarity	51.4%; Pred. No. 3.1e-34;	
Matches 110: Conservative	38; Mismatches	51;
Indels	15;	Gaps
3;		

Qy	60	1	DTPL	ETMY	ALRO	IVAS	GKAL	YVGI	SSYG	PEL	PAEA	EFMA	EGCP	LLIH	QPSY	SLIN	RW	60
Db	143	NTP	MEET	ASAL	HA	VO	SGKA	LYVG	ISYS	PSP	QKV	ELL	REWK	PILL	IHQ	PSY	SLIN	202
Qy	61	VEP	GD	GEN	LNS	OA	NN	GG	VTAF	SP	LA	QGLT	TDKY	LQGI	PEGR	---	ASQ	116
Db	203	VD	XSG	----	LLD	TQ	NN	VG	C	TA	FT	PL	QA	GLT	GT	YL	NG	257
Qy	117	ML	VN	IN	DM	YR	K	L	IN	DI	AE	QR	SG	LA	QMA	LAW	LE	176
Db	258	ML	TE	AN	LNS	R	L	L	N	EA	QQ	RQ	SG	QA	MA	L	SW	311
Qy	177	LD	SL	N	LF	SD	AE	E	LE	A	I	D	E	I	S	H	AG	210
Db	312	VO	AL	NN	L	T	F	T	K	EL	AO	I	D	O	H	I	AG	345

RESULT 2

probable reductase Z4354 [Imported] - Escherichia coli (strain O157:H7)
E83959
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 31-Mar-2001
C:Accession: E83959
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Fouts, D.E.;
Miller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Diallanta, E.; Potamoukou,
N.; and Salzberg, S.L. 2001.
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7
A:Reference number: A85480; PMID:21074935; PMID:11206551

A:Accession: E85959

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-346 <STO>

A:Cross-references: GB:AE005174; NID:g12517567; PIDN:AAG58137.1; GSPDB:GNO0145; UWGP:243

A:Experimental source: strain O157:H7, substrain EDL933

C:Genetics:

A:Gene: 24354

C:Superfamily: fission yeast pyridoxine 4-dehydrogenase

Query Match 48.3%; Score 526.5; DB 2; Length 346;

Best Local Similarity 50.9%; Pred. No. 5.3e-34;

Matches 109; Conservative 39; Mismatches 51; Indels 15; Gaps 3;

QY 1 DTPLEETMYALRDIVASGKALYVGISYGPDLTAEEAFMAEEGCPLLIHOPYSYIINRW 60

Db 143 NTPEETASALAHAVGSKALYVGISYSPERTOKKVELLHEWKIPLLIHOPYSYIINRW 202

QY 61 VEPPGDDGENLQSAANNGLGVIAFPLAQGLLTKYLDGIPESGR----ASQCKSLSEG 116

Db 203 VDKSG----LLDTLQNNVGCIATFPLAQGLLTKYLDGIPESGRMHREGNKVRLTPK 257

QY 117 MLVNNIDMYRKNDIAQERGQSLAQMALAWLRLRQEGYADTVTSALIGASSVQEQLDNS 176

Db 258 MLTEANLNSURLNEMAAQOQSGMAQMSWLLKQ-----ERTSVLVGASRAEQLEEN 311

QY 177 LDSLNLEFSDAELEAIDEISHDAGINIAKATD 210

Db 312 VQALNNLTFTEBLAQIDQIADGELNLWQASSD 345

RESULT 3

T35337

probable ion channel subunit - Streptomyces coelicolor

C:Species: Streptomyces coelicolor

C:Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 18-Feb-2000

C:Accession: T35337

R:Oliver, K.; Harris, D.; James, K.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.

A:Reference number: 221575

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-319 <OLI>

A:Cross-references: EMBL:AL049863; PIDN:CAB42946.1; GSPDB:GNO0070; SCOEDB:SC5H1.21c

A:Experimental source: strain A3(2)

C:Genetics:

A:Gene: SCOEDB:SC5H1.21c

C:Superfamily: fission yeast pyridoxine 4-dehydrogenase

Query Match 41.2%; Score 449; DB 2; Length 319;

Best Local Similarity 53.6%; Pred. No. 5.8e-28;

Matches 105; Conservative 25; Mismatches 54; Indels 12; Gaps 4;

QY 2 TPLEETMYALRDIVASGKALYVGISYGPDLTAEEAFMAEEGCPLLIHOPYSYIINRW 61

Db 132 TPLEETMGALHTAVQOQKALYGVSVSPQETREAAAILAELGTPLLIHQPRYSMLDRP 191

QY 62 EEPGDDGENLQSAANNGLGVIAFPLAQGLLTKYLDGIPESGR-ASQCKSLSEGMLNV 120

Db 192 ETGG-----LLDALDELEVGYSTAYSLEQGLLTGRYLNGIPEDSRAASDPFLNSDAVTE 246

QY 121 NNTDMYRKNDIAQERGQSLAQMALAWLRLRQEGYADTVTSALIGASSVQEQLDNSLDL 180

Db 247 ELVGKLRGLNEIATSGRQSLAQMALAWLRLRQEGYADTVTSALIGASSVQEQLEDSVAI 300

QY 181 NNLEFSDAELEAIDEI 196

Db 301 GNLDFADELARIDKI 316

RESULT 4

B86873

oxidoreductase yugB [Imported] - Lactococcus lactis subsp. lactis (strain IL1403)

C:Species: Lactococcus lactis subsp. lactis

C:Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 31-Mar-2001

C:Accession: B86873

R:Boitoutin, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malarme, K.; Weissenbach, J.; Eh

Genome Res. in press, 2001

A:Title: The complete genome sequence of the lactic acid bacterium.

A:Reference number: A86625

A:Accession: B86873

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-287 <STO>

A:Cross-references: GB:AE005176; NID:g12725031; PIDN:AAK06084.1; GSPDB:GNO0146

A:Experimental source: strain IL1403

C:Genetics:

A:Gene: yugB

C:Superfamily: fission yeast pyridoxine 4-dehydrogenase

Query Match 33.5%; Score 366; DB 2; Length 287;

Best Local Similarity 54.6%; Pred. No. 1.7e-21;

Matches 83; Conservative 22; Mismatches 37; Indels 10; Gaps 5;

QY 1 DTPLEETMYALRDIVASGKALYVGISYGPDLTAEEAFMAEE-GCPLLIHOPYSYIINR 59

Db 143 NTPIETMGALKSALDSGKALYVGLSNYSAKET-EAAVLAEEKLGHKLLIHQPRYSMLDR 201

QY 60 WVEPGDDGENLQSAANNGLGVIAFPLAQGLLTKYLDGIPESGRASOG--KSUSEGM 117

Db 202 WIE---DDLQETLLEA---GIGIAFKPLYQGLLTKYLDGIPESGRMRDPHYATLHDDS 255

QY 118 LNVNNDIMYRKNDIAQERGQSLAQMALAWVL 149

Db 256 LTKKRLQVQALNDLAQSRGQSLAQMALAWVL 287

RESULT 5

A75289

probable potassium channel, beta subunit - Deinococcus radiodurans (strain R1)

C:Species: Deinococcus radiodurans

C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000

C:Accession: A75289

R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J

M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.;

S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.

Science 286, 1571-1577, 1999

A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.

A:Reference number: A75250; MUID:20036896

A:Accession: A75289

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-315 <WHI>

A:Cross-references: GB:AE002063; GB:AE000513; NID:g6460121; PIDN:AAF11861.1; PID:g646

A:Experimental source: strain R1

C:Genetics:

A:Gene: DR2317

A:Map position: 1

C:Superfamily: fission yeast pyridoxine 4-dehydrogenase

Query Match 26.0%; Score 283.5; DB 2; Length 315;

Best Local Similarity 34.5%; Pred. No. 5.7e-15;

Matches 68; Conservative 40; Mismatches 76; Indels 13; Gaps 4;

QY 1 DTPLEETMYALRDIVASGKALYVGISYGPDLTAEEAFMAEEGCPPL-LIHOPYSYIINR 59

Db 127 EYPMEEIVMAFDQVTRDGKALYWGTSWMPAARIAQAVEFAKANGLHAPYTEQPEYSWVR 186

QY 60 WVEPGDDGENLQSAANNGLGVIAFPLAQGLLTKYLDGIPESGRASQCKSLSEGMLN 119

Db 187 DRVE-----QEILPYTEGAGIGLVVWSPLAMGLLTKYDEGREGARLTKENWAGSYLT 241

Qy	120 VNNIDMYRKLNDIAERGSQAQMALAWLREQGEYGADTVTSALIGASSVEQLDNLSDS	179
Dd	: : : : : : : : : : : : : : : : :	:
Dd	242 DENTIQRYDLKSIADDLGVTGTAQLALALLRKQG-----VSSVITGATKVNOIQT VKA	295
Qy	180 LNNLEFSDAELEAIDEI	196
Dd	: : : : : : : : : : : : : : : :	:
Dd	296 A-GVRLLDDVVORRIEDI	311
 RESULT 6		
H72391		
K+ channel, beta subunit - Thermotoga maritima (strain MSB8)		
C:Species: Thermotoga maritima		
C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000		
C:Accession: H72391		
R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey, Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.		
C.W.		
Nature 399, 323-329, 1999		
A;Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome se-		
A:Reference number: A72200; PMID:99287316		
A:Accession: H72391		
A>Status: preliminary		
A:Molecule type: DNA		
A:Residues: 1-319 <ARN>		
A:Cross-references: GB:AE001713; GB:AE000512; NID:g4980809; PIDN:AAD35401.1; PID:g498081		
A:Experimental source: strain MSB8		
C:Genetics:		
A:Gene: TW0313		
C:Superfamily: fission yeast pyrldoxine 4-dehydrogenase		

Query Match	24.5%	Score 267;	DB 2;	Length 319;
Best Local Similarity	33.8%	Pred. No. 1.le-13;		
Matches	71;	Conservative 43;	Mismatches 70;	Indels 26; Gaps 7;
Qy	1	DTPLEETMYALRDIVASGKALYVIGISSYGGPBLTAAEAEFMAEEG-CPLLIHOPSYSIINR	59	
Db	125	NVPMEVVFANDYILREGALYWGCTSEWSAKIEIEAHRVCKELGVMPPIVEQPYNNFVR	184	
Qy	60	WVEEPGDGGENLQSA-----NNGLGVIATSPLAQGLLTDKYLDGIEGSRASO-----GK	111	
Db	185	-----ERVEKEYAPLEYKMGITTSPLASGLLSGKYNNGIPEGSRLATFPQVRK	235	
Qy	112	SLSP-GMLNVNNIDMWKRLNDIAOERGOSLAQMALANVLREOGEYGADTVTSALIGASSV	170	
Db	236	WLEEGGLNKEKTFKKLRKLQNIADQLGASLPQLAIAWILNKK-----NVSSVILGVSRP	289	
Qy	171	EQLDNLSDLSNNLE-FSDAELEAIDEISHD	199	
Db	290	EQLEENIKAVEIKEKLTEDVMEETKEKILNE	319	

```

RESULT      7
T44988
oxidoreductase [imported] - Haloferax volcanii megaplasmid phv3
C:Species: Haloferax volcanii
C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 17-Mar-2000
C:Accession: T44988
R:Farahani, R.; Imbeault, J.C.; St Jean, A.; Chan, C.C.Y.; Allard, G.; Charlebois, R.L.
submitted to the EMBL Data Library, March 1997
A:Description: Hereditary instability of the megaplasmid phv3, and filamentation in the
A:Reference number: Z22886
A:Accession: T44988
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-336 <FAR>
A:Cross-references: EMBL:U95374; PIDN:AAB71807.1
A:Experimental source: strain DS2
C:Genetics:
A:Map position: megaplasmid phv3
A:Genome: plasmid

```

A;Note: expressed during exponential growth
C;Superfamily: fission yeast pyridoxine 4-dehydrogenase

Query Match	23.6%	Score 257;	DB 2;	Length 336;
Best Local Similarity	32.1%;	Pred. No. 7.5e-13;		
Matches 67;	Conservative 40;	Mismatches 84;	Indels 18;	Gaps 7
Qy	1	DTPLEETMYALRDIVASGKALYVVGISSSYGPEL--TAAEAEFMAEEGC-PLLIHQPYSYII	57	
		::::: :		
Db	141	ETPREMMKTNLNGVEDCKVHYLGASTLRPNNAWKVARANEIARAEGWEPFTVAQPRYNLV	200	
Qy	58	NRWVEEPGGDGENLLOSAANNGLGVIAESPQAQGLLTOKYL--DGIPGSRAQSQKSUSE	115	
		: :		
Db	201	DREIE--GD----YLEMTRSYCIAVPSPSQGFLTKYTREDGLTGESRAAESRFEE	254	
Qy	116	GMNVNNTDMVRKLNDAQERGQSILAOMALAWLRECEQEYGADTTVSALIGASSVEQLDN	175	
		: :		
Db	255	SYLTEENFDVHDELDVAGEVDATPAQTALAWLMURDG-----VTAPIVGARTVEQLTE	308	
Qy	176	SLDSLNINLEFSDAELEAIDEISHDAGINI	204	
		: :		
Db	309	NLEAA-TIDLTDQEVDRLTGAKPDPYVGL	336	

RESULT 8
C84256
probable oxidoreductase [imported] - Halobacterium sp. NRC-1
C:Species: Halobacterium sp. NRC-1
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 16-Feb-2001
C:Accession: C84256
R;Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.;
Leithausner, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.
Jung, K.H.; Alam, M.; Freitas, T.
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000
A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe,
A.; Title: Genome sequence of Halobacterium species NRC-1.

A:Residues: 1-336 <SVU>
A:Cross-references: GB:AE004437; NID:g10580553; PIDN:AAG19415.1; GSPDB:GN00138
C:Genetics:
A:Gene: yaj02
C:Superfamily: fission yeast pyridoxine 4-dehydrogenase

[illegible]

RESULT 9
F82815
voltage-ga
C:Species:

F82815 voltage-gated potassium channel beta subunit XF0367 [imported] - *Xylella fastidiosa* (C:Species: *Xylella fastidiosa*)

[illegible]

A:llid: Genome sequence of enterohemorrhagic *Escherichia coli* O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: E85538
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-324 <SPO>
A:Cross-references: GB:AE005174; NID:g12513275; PIDN:AAG54769.1; GSPDB:GN00145;
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: yajO
C:Superfamily: fission yeast pyridoxine 4-dehydrogenase

```

QY      60  WVEPGGDGNLLQSAANNGLGVIAFSPLAQGLLTDKYLDGIPGSRASQ-----GKSL 113
          : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      192  -----EEEREMPLCYOEGVAVIPQWSPPLARGLTR-----PWGETTARLVSEVGNKL 239
          : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      114  SEGLMNVNNDIMVRKLLNDIAQERQGSIAQMALAWVLREQGYGADTVTSALIGASSVEQL 173
          : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      240  YKES-DENDAOIAERLTGVSPEELGATRAQVALAWLLSKPG-----TAAPITGTSREQL 292
          : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      174  DNSLDSLNL 181
          | : : : :
Db      293  DELLNAVD 300

RESULT  12
C64771
probable oxidoreductase (EC 1.-.-.-) yajO - Escherichia coli
C:Species: Escherichia coli
C:Date: 12-Sep-1997 #sequence_revinsion 17-Sep-1997 #text_change 18-Feb-2000
C:Accession: C64771
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, A.; Rose, D.J.; Mau, B.; Shao, Y.

```



```

Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617
A:Accession: C64771
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-348 <BLAT>
A:Cross-references: GB:AF000148; GB:U000096; NID:g1786614; PIDN:AAC73522.1; PID:g1786621;
A:Experimental source: strain K-12, substrain MG1655
C:Genetics:
A:Gene: yajO
C:Superfamily: fission yeast pyridoxine 4-dehydrogenase
C:Keywords: oxidoreductase
F:151/Active site: His #status predicted

Query Match          20.6%; Score 225; DB 2; Length 348;
Best Local Similarity 31.9%; Pred. No. 2.6e-10;
Matches 60; Conservative 39; Mismatches 63; Indels 26; Gaps 6;

Qy 1 DTPLEETMYALRDIVASGKALYGVIGSSYPGLTAEAAEFMAEEGCPLIH-QPSYSIINR 59
      :|||::||:||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 156 NTPTEETLEALNDVVVKAGKARYIGASSMHASQFAQALELQKHGWAOQVFSQMDHYNLIYR 215
      ::|||::||:||:||||:||||:||||:||||:||||:||||:||||:||||:

Qy 60 WVEPEGDGENLLASANNGLVGITAFSPLAGCLLTDKYLDGIPEGSRAQ-----GKSL 113
      ::|||::||:||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 216 -----EEREEMPLCYOGEVAVIPWSPRLARGLTR-----PWGETTARLVSDVGVGNL 263
      ::|||::||:||:||||:||||:||||:||||:||||:||||:||||:||||:

Qy 114 SEGMLNVNNIDMVRKLNDIAOERGOSLAOMALANVLREOGEYGADTVTSALIGASSVEQL 173
      :|||::||:||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 264 YKES-DENDAQTAERLTCVSEELGATRAQVALALLSKPG-----IAAPIGTSTREEQL 316
      :|||::||:||:||||:||||:||||:||||:||||:||||:||||:||||:

Qy 174 DNSLDSLNL 181
      ||::|::
Db 317 DELLNAVD 324
      ||::|::

RESULT 13
T44934
moca protein [imported] - Agrobacterium tumefaciens plasmid pTi15955
C:Species: Agrobacterium tumefaciens
C>Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jul-2000
C:Accession: T44934
R:Kim, K.S.; Farrand, S.K.
J. Bacteriol. 178, 3275-3284, 1996
A:Title: Ti plasmid-encoded genes responsible for catabolism of the crown gall opine manure by the plant tumor.
A:Reference number: 222872; MUID:96236046
A:Accession: T44934
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-354 <KIM>
A:Cross-references: EMBL:U19620; NID:g797330; PIDN:AA807785.1; PFD:g797336
A:Experimental source: strain 15955
C:Genetics:
A:Gene: moca
A:Genome: plasmid pti15955
C:Superfamily: fission yeast pyridoxine 4-dehydrogenase

Query Match          19.5%; Score 213; DB 2; Length 354;
Best Local Similarity 32.5%; Pred. No. 2.3e-09;
Matches 69; Conservative 35; Mismatches 66; Indels 42; Gaps 9;

Qy 2 TPLEETMYALRDIVASGKALYGVIGSSY-----GPFLTAAEAFFMAEEGCPLLIIHQP 52
      |||::|||::||:||:||||:||||:||||:||||:||||:||||:||||:
Db 134 TPLEETMEALDTLRQGKVRYIGCSNETGTQIMKALGISEKDKRQRFVSQQ-----IH-- 186
      :|||::|||::||:||:||||:||||:||||:||||:||||:||||:||||:

Qy 53 SYSIINRWVEPPGDGENLLSANNGLVGITAFSPLAGCLLTDKYL--DGTPEGSRASQG 110
      :|||::|||::||:||:||||:||||:||||:||||:||||:||||:||||:
Db 187 -YTL-----EADAETVELPIPSVDQGLVLINSPLAGGLLSGKRRNQSAPEGSROFAG 239
      :|||::|||::||:||:||||:||||:||||:||||:||||:||||:||||:

Qy 111 KSL-----SEGMLNVNNIDMVRKLNDIAOERGOSLAOMALANVLREOGEYGADTVTSALI 165

```



```

Db      240  VSEBGRKQKNKLDLSPIAERLCTTLPQAVAWCLRNEG-----VSSVLLGSSTPEQLI 293
QY      175  NSLDSLNNL-EFSDAELEAIDEI 196
Db      294  ENLCATQVLPKMTSHVVNEIDNI 316

RESULT      3
US-08-606-143-2
; Sequence 2, Application US/08606143
; Patent No. 5856155
; GENERAL INFORMATION:
; APPLICANT: Li, Min

```

```

1  TITLE OF INVENTION: MODULATING POTASSIUM ION CHANNELS AND ASSAYS FOR
2  TITLE OF INVENTION: COMPOUNDS
3  NUMBER OF SEQUENCES: 45
4  CORRESPONDENCE ADDRESS:
5  ADDRESSEE: Leydig, Voit & Mayer, Ltd.
6  STREET: Two Prudential Plaza, Suite 4900
7  CITY: Chicago
8  STATE: IL
9  COUNTRY: USA
10 ZIP: 60601
11 COMPUTER READABLE FORM:
12 MEDIUM TYPE: Floppy disk
13 COMPUTER: IBM PC compatible
14 OPERATING SYSTEM: PC-DOS/MS-DOS
15 SOFTWARE: PatentIn Release #1.0, Version #1.25
16 CURRENT APPLICATION DATA:
17 APPLICATION NUMBER: US/08/606,143
18 FILING DATE: 23-FEB-1996
19 CLASSIFICATION: 514
20 ATTORNEY/AGENT INFORMATION:
21 NAME: Kilyk Jr., John
22 REGISTRATION NUMBER: 30763
23 REFERENCE/DOCKET NUMBER: 71756
24 TELECOMMUNICATION INFORMATION:
25 TELEPHONE: (312) 616-5600
26 TELEFAX: (312) 616-5700
27 TELEX: 25-3533
28 INFORMATION FOR SEQ ID NO: 2:
29 SEQUENCE CHARACTERISTICS:
30 LENGTH: 329 amino acids
31 TYPE: amino acid
32 STRANDEDNESS: single
33 TOPOLOGY: linear
34 MOLECULE TYPE: protein
35 US-08-606-143-2
36
37 Query Match 18.5%; Score 202; DB 2; Length 329;
38 Best Local Similarity 31.2%; Pred. No. 3.3e-14;
39 Matches 64; Conservative 37; Mismatches 82; Indels 22; Gaps
40
41 1 DTPLLEETWALRDIVASGKALVGVISSYGPELTSAAAEFMAEEG-CPLLHOPSYSIINR 59

```

```

QY      60 WVEPGDGENLQSANNGLGVIAFSPLAQGLLTDKYLDPGIESGRAS-----QGKS 112
        |         :         :         :         :         :         :         :
Db     185 EKVE-----VOLPELFHKTCVGAMTWSPCLACGVISGKYDGIPPPYSRASLKGYNWLKDKI 239
        |         :         :         :         :         :         :         :
QY     113 LSEGMNLNVNNIDMVRKLNDIAQERGOSLAQMALAWVLRQEGYGADVTVTSALIGASSVQE 172
        |||       :|||   :|||   :|||   :|||   :|||   :|||   :|||   :|||
Db    240 LSEE--GRQQAQLKELOAIERLCCTLPQLAIANCLRNEG-----VSSVLLGASNAEQ 291
                                     :
QY     173 LONSLDSLNL-EFSDAELEAIDEI 196
        |         :         :         :         :         :         :
Db    292 LAENIGAIVQLPKLSSSIVHEIDS I 316

```

US-08-576-626A-31
 ; Sequence 31, Application US/08576626A
 ; Patent No. 5998194
 ; GENERAL INFORMATION:
 ; APPLICANT: Summers, R.G.
 ; APPLICANT: Katz, L.
 ; APPLICANT: Donadio, S.
 ; APPLICANT: Staver, M.J.
 ; TITLE OF INVENTION: POLYKETIDE-ASSOCIATED SUGAR
 ; TITLE OF INVENTION: BIOSYNTHESIS GENES
 ; NUMBER OF SEQUENCES: 60
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Abbott Laboratories
 ; STREET: 100 Abbott Park Road
 ; CITY: Abbott Park
 ; STATE: Illinois
 ; COUNTRY: USA
 ; ZIP: 60064-3500
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette
 ; COMPUTER: IBM Compatible
 ; OPERATING SYSTEM: DOS
 ; SOFTWARE: FastSeq Version 2.0
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/576,626A
 ; FILING DATE: 21-DEC-1995
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER:
 ; FILING DATE:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Dianne Casuto
 ; REGISTRATION NUMBER: P-40,943
 ; REFERENCE/DOCKET NUMBER: 5857.US.O1
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (847) 938-3137
 ; TELEFAX: (847) 938-2623
 ; TELEX:
 ; INFORMATION FOR SEQ ID NO: 31:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1114 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: No. 5998194e
 ; US-08-576-626A-31

Query Match 16.0%; Score 175; DB 2; Length 1114;
 Best Local Similarity 29.9%; Pred. No. 2.3e-10;
 Matches 59; Conservative 38; Mismatches 80; Indels 20; Gaps 7;
 QY 3 PLEETMYALRDIVASGKALYVGISYGPGLTAAAEFMA-EEGCPLLIHOPSYIINRW 61
 Db 919 PWDEVWQMSLVASGKSYVGSNFGWIAAQAENARRHSLGWSHQCCLYNLVRHA 978
 QY 62 EEPGDDGNNLQSAANGLGVAFSPLAQGLLTDKYLDPGPEGS--RASOGKSLSEGMLN 119
 Db 979 E-----LEVLPAQAAYGLGVFAMSPHLGGLLSGA-LEKLAAGTAVKSAQGRA---QVLL 1028
 QY 120 VNNIDMYRKLNDAIQAERGOSIAQMALAWLREOGEYADTVTSALIGASSVEOLDNSLDS 179
 Db 1029 PSURPAEAEVKEFCRNGEDPAEVLAWLVRSPG-----IAGAVIGPRTPEQLDSALKA 1082
 QY 180 LNNLEFSDAELEAIDEI 196
 Db 1083 -SMTLDEQALSELDEI 1098

RESULT 5
 US-08-907-674-1
 ; Sequence 1, Application US/08907674
 ; Patent No. 5919685

; GENERAL INFORMATION:
 ; APPLICANT: Bandman, Olga
 ; APPLICANT: Shah, Purvi
 ; APPLICANT: Guegler, Karl J.
 ; APPLICANT: Corley, Neil C.
 ; TITLE OF INVENTION: HUMAN AFLATOXIN B1 ALDEHYDE REDUCTASE
 ; NUMBER OF SEQUENCES: 3
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Incyte Pharmaceuticals, Inc.
 ; STREET: 3174 Porter Drive
 ; CITY: Palo Alto
 ; STATE: CA
 ; COUNTRY: USA
 ; ZIP: 94304
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette
 ; COMPUTER: IBM Compatible
 ; OPERATING SYSTEM: DOS
 ; SOFTWARE: FastSeq for Windows Version 2.0
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/907,674
 ; FILING DATE: Herewith
 ; CLASSIFICATION: 514
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER:
 ; FILING DATE:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Billings, Lucy J.
 ; REGISTRATION NUMBER: 36,749
 ; REFERENCE/DOCKET NUMBER: PF-0362 US
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 415-855-0555
 ; TELEFAX: 415-845-4166
 ; TELEX:
 ; INFORMATION FOR SEQ ID NO: 1:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 331 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; IMMEDIATE SOURCE:
 ; LIBRARY: BRAINOT14
 ; CLONE: 1596452
 ; US-08-907-674-1

Query Match 14.9%; Score 162.5; DB 2; Length 331;
 Best Local Similarity 27.8%; Pred. No. 8e-10;
 Matches 59; Conservative 34; Mismatches 94; Indels 25; Gaps 7;
 QY 2 TPLEETMYALRDIVASGKALYVGISYGPGLTAAAEFMAEEGCPPL-LIHOPSYIINRW 60
 Db 119 TPVEETLHACORLHQECKFVELGSLNYSVAEVAEICTLCKSNGWILPTVYOGMYNAITRQ 178
 QY 61 VEPPGDDGNNLQSAANGLGVAFSPLAQGLLTDKYLDPGPEGSRASQ---GKSLSEG 117
 Db 179 VE-----TELPCLRFHGLRFYAYNPLAGLLTGKYEKDKGQKQPVGRFFGNTWAEY 232
 QY 118 LN-----VNNIDMYRKLNDAIQ-ERGOSIAQMALAWLRE---QGEYGADTVTSALIG 166
 Db 233 RNRVKEHHEGEGALVEKALQAAYGASAPSVTSAAALRWYHHSOLOGAHG-----DAVILG 288
 QY 167 ASSVEQLDNLSDLSNNLEFSDAELEAIDEISH 198
 Db 289 MSSLEQLEQNLAATEEGPLEPVAVVDFAPNQAWH 320

RESULT 6
 US-09-215-087-1
 ; Sequence 1, Application US/09215087
 ; Patent No. 5981244
 ; GENERAL INFORMATION:
 ; APPLICANT: Bandman, Olga

```

; APPLICANT: Shah, Purvi
; APPLICANT: Guegler, Karl J.
; APPLICANT: Corley, Neil C.
; TITLE OF INVENTION: HUMAN AFLATOXIN B1 ALDEHYDE REDUCTASE
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/215,087
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/907,674
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0362 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 331 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: BRAINOT14
; CLONE: 1596452
;
US-09-215-087-1

```

```

Query Match 14.9%; Score 162.5; DB 2; Length 331;
Best Local Similarity 27.8%; Pred. No. 8e-10;
Matches 59; Conservative 34; Mismatches 94; Indels 25; Gaps 7;

QY 2 TPLEETMYALRDIVASGKALYVGISSYGPETLTAEEAFNAEEGCPL-LIHQPSYSIINRW 60
DB 119 TPVEETLHACQRLHQBKFEVGLSLNYSASWEVAEICTLCKSNGLWILPTVYQGMYNATIRQ 178
QY 61 VEEPGDGGENLLQSAANGLGVITAFSPLAGGLTKYLDGIPEGRASQ---GKSLSEGM 117
DB 179 VE-----TELPCLRHFGLRYAYNPLAGGLTKYKEDKDGKPGVGRFGNTWAEMY 232
QY 118 LN-----VNNIDMVRLNDIAQ-ERGOSLAQMALAWLRE---QGEYGADTVTSALIG 166
DB 233 RNRVKEHHEFGIALVEKALQAAYGASPSVTSALRWMYHHSQLOQAHG----DAVILG 288
QY 167 ASSVEOLDNSLDLNNLEFSDAELEAIDEISH 198
DB 289 MSSLEQLEQNLAATEEGPLEPAVVDFAFNQAWH 320

```

```

RESULT 7
US-09-391-959-1
; Sequence 1, Application US/09391959
; Patent No. 6071704
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Shah, Purvi
; APPLICANT: Guegler, Karl J.

```

```

; APPLICANT: Corley, Neil C.
; TITLE OF INVENTION: HUMAN AFLATOXIN B1 ALDEHYDE REDUCTASE
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/391,959
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/907,674
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0362 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 331 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: BRAINOT14
; CLONE: 1596452
;
US-09-391-959-1

```

```

Query Match 14.9%; Score 162.5; DB 3; Length 331;
Best Local Similarity 27.8%; Pred. No. 8e-10;
Matches 59; Conservative 34; Mismatches 94; Indels 25; Gaps 7;

QY 2 TPLEETMYALRDIVASGKALYVGISSYGPETLTAEEAFNAEEGCPL-LIHQPSYSIINRW 60
DB 119 TPVEETLHACQRLHQBKFEVGLSLNYSASWEVAEICTLCKSNGLWILPTVYQGMYNATIRQ 178
QY 61 VEEPGDGGENLLQSAANGLGVITAFSPLAGGLTKYLDGIPEGRASQ---GKSLSEGM 117
DB 179 VE-----TELPCLRHFGLRYAYNPLAGGLTKYKEDKDGKPGVGRFGNTWAEMY 232
QY 118 LN-----VNNIDMVRLNDIAQ-ERGOSLAQMALAWLRE---QGEYGADTVTSALIG 166
DB 233 RNRVKEHHEFGIALVEKALQAAYGASPSVTSALRWMYHHSQLOQAHG----DAVILG 288
QY 167 ASSVEOLDNSLDLNNLEFSDAELEAIDEISH 198
DB 289 MSSLEQLEQNLAATEEGPLEPAVVDFAFNQAWH 320

```

```

RESULT 8
US-08-907-674-3
; Sequence 3, Application US/08907674
; Patent No. 5919685
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Shah, Purvi
; APPLICANT: Guegler, Karl J.
; APPLICANT: Corley, Neil C.
; TITLE OF INVENTION: HUMAN AFLATOXIN B1 ALDEHYDE REDUCTASE

```


CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/391,959
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/907,674
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0362 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 327 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: GenBank
CLONE: 433611
US-09-391-959-3

Query Match 14.3%; Score 156.5; DB 3; Length 327;
Best Local Similarity 26.3%; Pred. No. 3.6e-09;
Matches 57; Conservative 37; Mismatches 94; Indels 29; Gaps 8;
QY 2 TPLEETMYALRDIVASGKALYGVSSYGPETLTAEEAEFMAEEGCPL-LIHQPSYSIINRW 60
Db 115 TPLEETLQACHVHQEGKFGVLSNYSWEVAEITLCKKNGWIMPTVYQGMYNITRQ 174
QY 61 VERPDDGGENLQSAANNGIVAFSPLAQGLLTDKYLDCIPEGRASQ---CKSLSEG 117
Db 175 VE-----TELPCLRHFGRLFFAFNPLAGGLTGRYKQYQDKGKNPESRFFGNPESQLY 228
QY 118 L-----NVNNTDMVRK-LNDIAQERGOSLAOMALAWLRE---QGEYGADVTTSALIG 166
Db 229 MDYRWKEEHFNGLAVEKALKTTTGPTAPSMISAARVMYHHSQKGTQG----DAVILG 284
QY 167 ASSVEQLDNLSDLNLEFSDAELEAIDE---ISHD 199
Db 285 MSSLEQLQNLALVEEGPLEPAVVDADFQANLVAHE 321

RESULT 11
US-09-166-412-4
Sequence 4, Application US/09166412
Patent No. 6133504
GENERAL INFORMATION:
APPLICANT: Wayne D. Loescher,
APPLICANT: John D. Everard
APPLICANT: Rebecca Grunet
TITLE OF INVENTION: DNA Encoding Mannose
TITLE OF INVENTION: 6-Phosphate Reductase
TITLE OF INVENTION: and Recombinants
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Ian C. McLeod
STREET: 2190 Commons Parkway

CITY: Okemos
STATE: Michigan
COUNTRY: USA
ZIP: 48864
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette 5.25 inch,
MEDIUM TYPE: 360 kb storage
COMPUTER: IBM Compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: Wordperfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/166,412
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/731,320
FILING DATE: October 15, 1996
ATTORNEY/AGENT INFORMATION:
NAME: Ian C. McLeod
REGISTRATION NUMBER: 20,931
REFERENCE/DOCKET NUMBER: MSU 4.1-275
TELECOMMUNICATION INFORMATION:
TELEPHONE: (517) 347-4100
TELEFAX: (517) 347-4103
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 310
TYPE: Amino Acid
STRANDEDNESS: Single
TOPOLOGY: Linear
MOLECULE TYPE:
DESCRIPTION:
HYPOTHETICAL: No
ANTI-SENSE: No
ORIGINAL SOURCE:
ORGANISM: apple
IMMEDIATE SOURCE:
LIBRARY: N/A
US-09-166-412-4
Query Match 9.9%; Score 107.5; DB 4; Length 310;
Best Local Similarity 20.8%; Pred. No. 0.00091;
Matches 44; Conservative 47; Mismatches 72; Indels 49; Gaps 7;
QY 4 LEETMYALRDIVASGKALYGVSSYGPETLTAEEAEFMAEEGCPL-LIHQPSYSIINRWEE 63
Db 139 LOOTWEGMEKTVSLGLVRSIGLSNYELFUTRDC-----LAYSKIKPAVSQFETH 187
QY 64 PGDDGNNLQSAANNGIVAFSPLAQGLLTDKYLDCIPEGRASQKSLSEGMLNVNNI 123
Db 188 PYFQDRLVKFCMKHGVLTHTAHTPLG-GAANKDMFG-----SVSPL 228
QY 124 DMVRKLNIDIAQERGOSLAOMALAWLREGEYGAADVTTSALIGASVEQLDNLSDLSLNL 183
Db 229 D-DPVLNDVAKKYKGSVAQICLRW-----GIQKRTAVIPKSKIKRIQLENLEVL-EF 278
QY 184 EFSDAELEAIDEI-----SHDAGINIWA 206
Db 279 QLSDEDMQLIYSIDRKRYRTSLPSKWTGLDYYA 310

RESULT 12
US-09-166-412-2
Sequence 2, Application US/09166412
Patent No. 6133504
GENERAL INFORMATION:
APPLICANT: Wayne D. Loescher,
APPLICANT: John D. Everard
APPLICANT: Rebecca Grunet
TITLE OF INVENTION: DNA Encoding Mannose
TITLE OF INVENTION: 6-Phosphate Reductase

RESULT 13
US-08-336-198C-3
; Sequence 3, Application US/08336198C
; Patent No. 5866382

QY	1	DTPLEETWYALRDIVASGRALYGV	ISSYGP	ELTAAEAEAE	EGCPLLI	-HOPSYSLINR	59		
DB	140	DVPILETWALEKLVRAGIR	SIGVS	NFFGALLDL	DLRGATIKPSV	QVQVEHHP	192		
QY	60	WVEPDDCGENILLOSANNG	LVIAF	SPLAQGLL	TDKYL	DGPIEGSRASQGSLS	EGMLN 119		
DB	193	YLOQP----	RLIEFAQSGIAVT	YSSP-----	GPQS	FVELNQGRALNT	SPLF 236		
QY	120	VNNIDMVRKLNDIAQBRG	QSQAQMA	LAWLVLR	EQGEYG	DATVT	SALIGASSV	EQOLD -NSLD 178	
DB	237	EN-----	ETIKALAAKHG	SPQVLLRW-----	SSQ	RGIAIIPKSN	TVPRLL	ENKDNV	SFD 287
QY	179	SUNLEFSD-AELE	191						
DB	288	-LDEODFADIAKL	300						

RESULT 14
US-09-222-817-12
; Sequence 12, Application US/09222817
; Patent No. 6037154
; GENERAL INFORMATION:
; APPLICANT: Mikiko SUGA, Masatazu S
; TITLE OF INVENTION: METHOD OF PRODUCING
; FILE REFERENCE: 0P813
; CURRENT APPLICATION NUMBER: US/09/222-817-12
; CURRENT FILING DATE: 1998-12-30

Query Match	8.8%	Score 95.5;	DB 3;	Length 530;
Best Local Similarity	23.6%;	Pred. No. 0.047;		
Matches 59;	Conservative 30;	Mismatches 106;	Indels 55;	Gaps 9;

QY	10	ALRDIVASGKALYVGISSYGPETATAAEAPMAEEGCPLLIHOPSYSIINRWVEEPG--DD	67
		: : : : : :	
Db	243	ALADAIESGHIRGAGFDVSTPCTDSPLFK-----LPQVVVTPHLGASTAEAAQDRAGTDV	298
		: : : : : :	
QY	68	GENLLQSANN-----GLGVTAFSPLAQ--GLLTDKYLIDIGEPSRASQSKS	112
		: : : : : : : : : : :	
Db	299	ADSVLKALAGFEFVADAVNSGGRVGRKVAWMDLARKLGLLAGKLVDAAPVIEVEARGE	358
		: : : : :	

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 13, 2001, 17:51:44 ; Search time 34.57 Seconds
(without alignments)
508.561 Million cell updates/sec

Title: US-09-105-117i-3

Perfect score: 1460

Sequence: 1 MNPIDTLISIDEGSFG.....RSLRLTDAVDAIEGLRP 290

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 412676 seqs, 60623988 residues

Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :	A_Geneseq_0601.*
1:	/SID88/gcgdata/geneseq/geneseq/AA1980.DAT.*
2:	/SID88/gcgdata/geneseq/geneseq/AA1981.DAT.*
3:	/SID88/gcgdata/geneseq/geneseq/AA1982.DAT.*
4:	/SID88/gcgdata/geneseq/geneseq/AA1983.DAT.*
5:	/SID88/gcgdata/geneseq/geneseq/AA1984.DAT.*
6:	/SID88/gcgdata/geneseq/geneseq/AA1985.DAT.*
7:	/SID88/gcgdata/geneseq/geneseq/AA1986.DAT.*
8:	/SID88/gcgdata/geneseq/geneseq/AA1987.DAT.*
9:	/SID88/gcgdata/geneseq/geneseq/AA1988.DAT.*
10:	/SID88/gcgdata/geneseq/geneseq/AA1989.DAT.*
11:	/SID88/gcgdata/geneseq/geneseq/AA1990.DAT.*
12:	/SID88/gcgdata/geneseq/geneseq/AA1991.DAT.*
13:	/SID88/gcgdata/geneseq/geneseq/AA1992.DAT.*
14:	/SID88/gcgdata/geneseq/geneseq/AA1993.DAT.*
15:	/SID88/gcgdata/geneseq/geneseq/AA1994.DAT.*
16:	/SID88/gcgdata/geneseq/geneseq/AA1995.DAT.*
17:	/SID88/gcgdata/geneseq/geneseq/AA1996.DAT.*
18:	/SID88/gcgdata/geneseq/geneseq/AA1997.DAT.*
19:	/SID88/gcgdata/geneseq/geneseq/AA1998.DAT.*
20:	/SID88/gcgdata/geneseq/geneseq/AA1999.DAT.*
21:	/SID88/gcgdata/geneseq/geneseq/AA2000.DAT.*
22:	/SID88/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1460	100.0	290	18	AAW37714 C. glutamicum Lys
2	1460	100.0	290	22	AAW79658 Corynebacterium gl
3	118.5	8.1	278	20	AAV39317 ORF116 protein inv
4	105	7.2	311	19	AAW36129 Snpr activator pro
5	105	7.2	311	21	AAW30494 Amino acid sequenc
6	102.5	7.0	299	21	AAV74592 Neisseria meningit
7	101.5	7.0	299	21	AAV74593 Neisseria meningit
8	100.5	6.9	344	20	AAV29277 Protein encoded by
9	99.5	6.8	1346	21	AAW18640 Amino acid sequenc
10	99.5	6.8	1346	21	AAV77195 S. venezuelae macr
11	99.5	6.8	1346	21	AAV67204 Narbonolide synth

12	99.5	6.8	12199	21	AAV77180
13	98	6.7	303	21	AAW74591
14	95.5	6.5	5069	19	AAW52846
15	93.5	6.4	298	13	AAW25588
16	92.5	6.3	928	21	AAV77292
17	92.5	6.3	928	21	AAV78844
18	92.5	6.3	4150	21	AAV92707
19	92	6.3	401	20	AAV33264
20	92	6.3	401	20	AAV33268
21	92	6.3	401	20	AAV33270
22	92	6.3	401	20	AAW92938
23	92	6.3	401	20	AAW92936
24	92	6.3	401	20	AAW92934
25	89.5	6.1	383	15	AAV49827
26	89.5	6.1	1346	21	AAV77203
27	89	6.1	290	17	AAW02378
28	87.5	6.0	432	16	AAW77865
29	87.5	6.0	3798	21	AAV58577
30	87	6.0	303	21	AAW80094
31	86.5	5.9	461	17	AAW98342
32	86.5	5.9	1319	14	AAW5002
33	86	5.9	391	21	AAV95652
34	84.5	5.8	3266	21	AAW42491
35	84.5	5.8	3739	21	AAW18638
36	84.5	5.8	3739	21	AAV77193
37	84.5	5.8	3739	21	AAV77201
38	84.5	5.8	3739	21	AAV67202
39	83.5	5.7	2188	22	AAW66467
40	83.5	5.7	3816	21	AAV92708
41	83	5.7	419	12	AAW10691
42	83	5.7	856	17	AAW96949
43	83	5.7	1688	19	AAW52848
44	82.5	5.7	2152	20	AAV39298
45	82.5	5.7	4472	18	AAW22601

ALIGNMENTS

RESULT	1
AAW37714	
ID	AAW37714 standard; Protein; 290 AA.
XX	
AC	AAW37714;
XX	
DT	12-MAR-1998 (first entry)
XX	
DE	C. glutamicum Lys G protein (lysine export regulator).
XX	
KW	LysG; LysE; ORF3; lysine transport; regulatory protein; export protein;
KW	Microbial production; amino acid; animal feed additive.
XX	
OS	Corynebacterium glutamicum.
XX	
FN	DE19548222-AL.
XX	
PD	26-JUN-1997.
XX	
PF	22-DEC-1995; 95DE-1048222.
XX	
PR	22-DEC-1995; 95DE-1048222.
XX	

(KERJ) FORSCHUNGSZENTRUM JUELICH GMBH.
Eggeling L, Sahm H, Vrljic M;
WPI: 1997-333867/31.
N-PSDB; AAT96816.
Increasing microbial production of amino acids, especially lysine -
by improving export carrier activity or corresponding gene
expression, also new export and regulatory genes from
Corynebacterium

XX Disclosure; Page 6; 16pp; German.

XX This sequence is the LysE protein product, a lysine export regulator.

XX LysG and LysE encode a lysine transport regulatory protein and an export

CC protein, respectively. Microbial production of amino acids (A) is

CC improved by increasing the export-carrier activity and/or the export gene

CC expression in a microorganism that produces (A). The method is

CC specifically used to increase production of lysine, used as an animal

CC feed additive. Other (A) are variously useful as pharmaceuticals,

CC condiments and intermediates for fine chemicals. This method increases

CC the amount of (A) secreted into the culture medium. Export of (A) has

CC been found to depend on a single gene.

XX Sequence 290 AA;

Query Match 100.0%; Score 1460; DB 18; Length 290;

Best Local Similarity 100.0%; Pred. No. 3.2e-150;

Matches 290; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNPIQDTLLSIIDEGSFEGASLALSTSPSAVSQVRKALEHHVGRVIVSTQPAKATEAG 60

Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 61 EVLVQARKMVLQAEKAGLQSLAEIPLTIAINADSLSTWFPFVNEVASWGATLTL 120

Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 121 RLEDEAHTLSLLRGDVLGAVTREPANPVAGCEVVEIGTMRHLAIAITPSLRDAYMVGKLD 180

Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 181 WAAMPVLRFGPKDVLQDRDLGVRDGPVGRRRYSIVPSAEGFEATRGGLGWLLPETQA 240

Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 241 APLMKAGEVILLDEIPIDTPMYQWRLESRLARLTDAVDDAIEGLRP 290

Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 2

AA879658

ID AAB79658 standard; Protein; 290 AA.

XX

AC AAB79658;

XX

DT 30-APR-2001 (first entry)

XX

DE Corynebacterium glutamicum MP protein sequence SEQ ID NO:50.

XX

KW Corynebacterium glutamicum; metabolic pathway protein; MP protein;

KW fine chemical production; microorganism; organic acid; nucleoside;

KW nonproteinogenic amino acid; purine base; pyrimidine base; nucleotide;

KW lipid; saturated fatty acid; unsaturated fatty acid; diol; vitamin;

KW carbohydrate; aromatic compound; cofactor; polyketide; enzyme.

XX

OS Corynebacterium glutamicum.

XX

PN W0200100843-A2.

XX

PD 04-JAN-2001.

XX

PF 23-JUN-2000; 2000WO-IB00923.

XX

PR 25-JUN-1999; 99US-0141031.

PR 01-JUL-1999; 99DE-1030476.

PR 02-JUL-1999; 99US-0142101.

PR 08-JUL-1999; 99DE-1031415.

PR 08-JUL-1999; 99DE-1031418.

PR 08-JUL-1999; 99DE-1031419.

PR 08-JUL-1999; 99DE-1031420.

PR

XX

08-JUL-1999; 99DE-1031424.

08-JUL-1999; 99DE-1031428.

08-JUL-1999; 99DE-1031434.

08-JUL-1999; 99DE-1031435.

08-JUL-1999; 99DE-1031443.

08-JUL-1999; 99DE-1031453.

08-JUL-1999; 99DE-1031457.

08-JUL-1999; 99DE-1031465.

08-JUL-1999; 99DE-1031478.

08-JUL-1999; 99DE-1031510.

08-JUL-1999; 99DE-1031541.

08-JUL-1999; 99DE-1031573.

08-JUL-1999; 99DE-1031592.

08-JUL-1999; 99DE-1031632.

08-JUL-1999; 99DE-1031634.

08-JUL-1999; 99DE-1031636.

09-JUL-1999; 99DE-1032125.

09-JUL-1999; 99DE-1032126.

09-JUL-1999; 99DE-1032130.

09-JUL-1999; 99DE-1032186.

09-JUL-1999; 99DE-1032206.

09-JUL-1999; 99DE-1032227.

09-JUL-1999; 99DE-1032228.

09-JUL-1999; 99DE-1032229.

09-JUL-1999; 99DE-1032230.

14-JUL-1999; 99DE-1032922.

14-JUL-1999; 99DE-1032926.

14-JUL-1999; 99DE-1032928.

14-JUL-1999; 99DE-1033004.

14-JUL-1999; 99DE-1033005.

14-JUL-1999; 99DE-1033006.

12-AUG-1999; 99US-0148613.

27-AUG-1999; 99DE-1040764.

27-AUG-1999; 99DE-1040765.

27-AUG-1999; 99DE-1040766.

27-AUG-1999; 99DE-1040832.

31-AUG-1999; 99DE-1041378.

31-AUG-1999; 99DE-1041379.

31-AUG-1999; 99DE-1041380.

31-AUG-1999; 99DE-1041394.

31-AUG-1999; 99DE-1041396.

03-SEP-1999; 99DE-1042076.

03-SEP-1999; 99DE-1042077.

03-SEP-1999; 99DE-1042079.

03-SEP-1999; 99DE-1042086.

03-SEP-1999; 99DE-1042087.

03-SEP-1999; 99DE-1042088.

03-SEP-1999; 99DE-1042095.

03-SEP-1999; 99DE-1042124.

03-SEP-1999; 99DE-1042129.

09-MAR-2000; 2000US-0187970.

XX

PA (BADI) BASF AG.

XX

PI Pompejus M, Kroeger B, Schroeder H, Zelder O, Haberhauer G;

XX

DR WPI; 2001-137957/14.

DR N-PSDB; AAF71777.

XX

PT Nucleic acids from Corynebacterium glutamicum encoding metabolic

PT pathway proteins, useful for producing fine chemicals in

PT microorganisms, including organic acids, nonproteinogenic amino acids,

PT and purine and pyrimidine bases -

XX

PS Claim 20; Page 228-229; 1737pp; English.

XX

CC AAF71753 to AAF72330 encode the Corynebacterium glutamicum metabolic

CC pathway (MP) proteins given in AAB79634 to AAB80211. The C. glutamicum

CC MP nucleic acids are useful for the production of fine chemicals

CC in microorganisms, including organic acids, nonproteinogenic amino

CC acids, purine and pyrimidine bases, nucleosides, nucleotides, lipids,

CC saturated and unsaturated fatty acids, diols, carbohydrates, aromatic

CC compounds, vitamins, cofactors, polyketides and enzymes.

[illegible]

XX PN WO9744439-A2.
 XX PD 27-NOV-1997.
 XX PF 22-MAY-1997; 97WO-US08690.
 XX PR 24-MAY-1996; 96US-0653650.
 XX PA (OHIS) UNIV OHIO STATE. RES FOUND.
 XX PI DeSanti CL, Dickens ML, Strohl WA;
 XX DR WPI; 1998-018495/02.
 XX DR N-PSDB; AAV01451.
 XX PT Producing doxorubicin from daunomycin by daunomycin C14 hydroxylase
 XX PT - also hydroxylation and oxidation of other anthracycline(s) with
 XX PT the same enzyme
 XX PS Disclosure; Fig 6; 59pp; English.
 XX CC This protein is the encoded product of the Streptomyces snpr
 XX CC activator gene. The snpr activator gene is incorporated in
 XX CC novel plasmid pANT195 (see AAV01451) that also includes the
 XX CC Snpr-activated snpA promoter and a modified doxA gene (see
 XX CC AAV01447) of Streptomyces sp. strain C5. The doxA gene codes for
 XX CC daunomycin C-14 hydroxylase (see AAW36128), an enzyme capable of
 XX CC converting daunomycin to the anticancer agent doxorubicin. Host
 XX CC cells, especially Streptomyces host cells, transformed with pANT195
 XX CC can be used in methods for the production of doxorubicin from
 XX CC daunomycin or for the hydroxylation and oxidation of other
 XX CC anthracyclines.
 XX SQ Sequence 311 AA;

Query Match 7.2%; Score 105; DB 19; Length 311;
 Best Local Similarity 21.2%; Pred. No. 0.0052;
 Matches 66; Conservative 50; Mismatches 129; Indels 66; Gaps 13;
 QY 6 LDTLLSTIDGSEFEGASLALISPSAVSORVKALEHHVGRVLSRTQP-AKATEAGEVLV 64
 Db 8 lralcaiaatgslhraarqlgtqpslstqlrriehelggalfvrtartcgrptplgrlvi 67
 QY 65 QAARKMWL-----LQATKAKLGRGLAEIPLTIAINAD-SLSTWFFPPVFNEVASWGATLT 119
 Db 68 srarplvaelcslvseara---aavadaslrvgstasralagw----lrlrhwqptlh 120
 QY 120 LRLEDEAHTLSLLRRGDVLGAVTREPANPVAGCEVVELGTMRHLAIATPRLDAYMVDGKL 179
 Db 121 mdvsana-llrmvadghldvafvheve-----gsir-----vpeglrvrlvqrep 165
 QY 180 DWAAAMPVLRFGPKDVLQDRDLGRVDPVGR-----RRVSIVPSAEGFGGAIRR 228
 Db 166 qfvcpl-----adhpaaeatasyaptwttrmidptvdgwnavrr 207
 QY 229 -----GLGWGLL--PETQAAPMLKAGEVILLDE--IPIDTPMWQWRLESRLATDA 279
 Db 208 vlraegldsrilhgdyhsaaylvatgevvtvqptspraetaavrllhgdplgvrillaa 267
 QY 280 VDAATEGLRP 290
 Db 268 rtdtelegvyp 278
 RESULT 5
 AAB30494
 ID AAB30494 standard; Protein; 311 AA.
 XX AAB30494;
 XX 06-MAR-2001 (first entry)

XX DE Amino acid sequence of Snpr encoded by plasmid pANT1201.
 XX KW Streptomyces sp. strain C5; SnpA; S. venezuelae; alpha-amylase;
 XX KW endostatin; cancer; tumour growth; angiogenesis.
 XX OS Streptomyces sp.
 XX PN WO200060945-A1.
 XX PD 19-OCT-2000.
 XX PF 12-APR-2000; 2000WO-US09747.
 XX PR 13-APR-1999; 99US-0129084.
 XX PA (MERI) MERCK & CO INC.
 XX PI Desanti CL, Strohl WR;
 XX DR WPI; 2000-686970/67.
 XX DR N-PSDB; AAC62024.
 XX PT Preparation of soluble recombinant endostatin involves transforming
 XX PT Streptomyces host with expression vector comprising nucleotide
 XX PT sequence encoding endostatin operably linked to linker and leader
 XX PT peptide
 XX PS Example 1; Fig 9A-B; 57pp; English.
 XX CC The present sequence is encoded by a fragment of pANT1201. This plasmid
 XX CC was used as a source of snpr, snpA promoter, and multiple cloning site.
 XX CC The specification describes a method for the production of soluble,
 XX CC recombinant human endostatin in Streptomyces. Leader sequences of
 XX CC Streptomyces sp. strain C5 SnpA and S. venezuelae alpha-amylase proteins
 XX CC are linked to the N-terminal of endostatin. This ensures that endostatin
 XX CC protein is produced as a secreted, soluble protein which needs no
 XX CC refolding, is stable in the fermentation broth and is produced in large
 XX CC quantities. The method is used for preparing soluble recombinant human,
 XX CC murine or primate endostatin, which is useful in the treatment of cancer,
 XX CC inhibition of tumour growth, inhibition of angiogenesis, isolation of
 XX CC receptors for endostatin and for identification of anti-angiogenic
 XX CC compounds in assays. The endostatin protein is produced as a secreted,
 XX CC soluble protein which needs no refolding, is stable in the fermentation
 XX CC broth and is produced in large quantities. Streptomyces are amenable
 XX CC for cultivation in large fermentations allowing for large quantities of
 XX CC soluble endostatin to be produced.
 XX SQ Sequence 311 AA;

Query Match 7.2%; Score 105; DB 21; Length 311;
 Best Local Similarity 21.2%; Pred. No. 0.0052;
 Matches 66; Conservative 50; Mismatches 129; Indels 66; Gaps 13;
 QY 6 LDTLLSTIDGSEFEGASLALISPSAVSORVKALEHHVGRVLSRTQP-AKATEAGEVLV 64
 Db 8 lralcaiaatgslhraarqlgtqpslstqlrriehelggalfvrtartcgrptplgrlvi 67
 QY 65 QAARKMWL-----LQATKAKLGRGLAEIPLTIAINAD-SLSTWFFPPVFNEVASWGATLT 119
 Db 68 srarplvaelcslvseara---aavadaslrvgstasralagw----lrlrhwqptlh 120
 QY 120 LRLEDEAHTLSLLRRGDVLGAVTREPANPVAGCEVVELGTMRHLAIATPRLDAYMVDGKL 179
 Db 121 mdvsana-llrmvadghldvafvheve-----gsir-----vpeglrvrlvqrep 165
 QY 180 DWAAAMPVLRFGPKDVLQDRDLGRVDPVGR-----RRVSIVPSAEGFGGAIRR 228
 Db 166 qfvcpl-----adhpaaeatasyaptwttrmidptvdgwnavrr 207
 QY 229 -----GLGWGLL--PETQAAPMLKAGEVILLDE--IPIDTPMWQWRLESRLATDA 279
 Db 208 vlraegldsrilhgdyhsaaylvatgevvtvqptspraetaavrllhgdplgvrillaa 267
 QY 280 VDAATEGLRP 290
 Db 268 rtdtelegvyp 278

Oy	2	NPIQLDTLLSIIDEGSFEGASLALSSPSAVSQVRKALEHHVGRVLVSR-TOPAKATEAG	60
I	:	: :	
Dd	4	nseeltvfvqvesgsfstaacqlamansavsrikrleeklgvnlrrtrqisltteeg	63
Oy	61	EVLVOARKVKVLQAEATKAQLSRLAEIPITTAINDSLSTWFPVP-----FNE-	109
I	:	: :	
Dd	64	aqyfrraqrlqemaadetmla-vheibpgvl----svdsampmvllhlaplaaktner	118
Oy	110	-----VASWGGATLTLRLEDEAHTLSLLRRGDVLGAVTREANPVAGCEVVELGTMRH	161
I	:	: :	
Dd	119	yphirlslvsedyinlierkvdia-----lragel-----ddsglrarhlfdorf	164
Oy	162	LAIATP 167	
I	:	: :	
Dd	165	rviasp 170	
RESULT	7		
AAY74593			
ID	AAY74593	standard; Protein; 299 AA.	
XX			
AC	AAY74593;		
XX			
DT	21-MAR-2000	(first entry)	
XX			
DE	Neisseria meningitidis	ORF 158 protein sequence SEQ ID NO:660.	
XX			
KW	Neisseria meningitidis;	Neisseria gonorrhoeae; antigen; vaccine;	
KW	antigenic; diagnosis;	immunogenic; infection; meningitis; septicaemia;	
KW	antibacterial; gene therapy.		
XX			
OS	Neisseria meningitidis.		
XX			
PN	WO9957280-A2.		
XX			
PD	11-NOV-1999.		
XX			
PF	30-APR-1999;	99WO-US09346.	
XX			
PR	01-MAY-1998;	98US-0083758.	
PR	31-JUL-1998;	98US-0094869.	
PR	02-SEP-1998;	98US-0098994.	
PR	02-SEP-1998;	98US-0099062.	
PR	09-OCT-1998;	98US-0103749.	
PR	09-OCT-1998;	98US-0103794.	
PR	09-OCT-1998;	98US-0103796.	
PR	25-FEB-1999;	99US-0121528.	
XX			
PA	(CHIR) CHIRON CORP.		
PA	(GENO.) INST GENOMIC RES.		
XX			
PI	Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;		
PI	Petersen J, Piazza M, Rappuoli R, Ratti G, Scaliato E, Scarselli M;		
PI	Tettelin H, Venter JC;		
XX			
DR	WPI: 2000-062150/05.		
DR	N-PSDB; AA253355.		
XX			
PT	Novel Neisserial polypeptides predicted to be useful antigens for		
PT	vaccines and diagnostics -		
XX			
PS	Claim 2; Page 453; 1453pp; English.		
XX			
CC	AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941		
CC	represent novel <i>Neisseria meningitidis</i> and <i>N. gonorrhoeae</i> polynucleotides		
CC	and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ5473 represent		
CC	PCR primers used in the exemplification of the present invention. The		
CC	polypeptides, the polynucleotides, antibodies and compositions of		
CC	the invention can be used as vaccines, as diagnostic reagents, and as		
CC	immunogenic compositions. The polypeptides can be used in the		
CC	manufacture of medicaments for treating or preventing infection due to		
CC	<i>Neisserial bacteria</i> (e.g. meningitis and septicaemia), to detect the		

Query Match	7.0%;	Score 102.5;	DB 21;	Length 299;
Best Local Similarity	26.3%;	Pred. NO. 0.0092;		
Matches 49;	Conservative 31;	Mismatches 67;	Indels 39;	Gaps 7;

PT New recombinant pick hydroxylase gene of Streptomyces venezuelae useful
 PT for converting ketolides to antibiotics and as antibiotics and
 PT intermediates in the synthesis of compounds with pharmaceutical value

XX Disclosure; Columns 13-14; 117pp; English.

XX The present sequence represents a narbonolide synthase subunit 4
 CC (PikAIV). The nucleotide sequence encoding it is used in the course of
 CC the invention. The specification describes a recombinant DNA compound
 CC expressing recombinant polyketide synthase genes in host cells for the
 CC production of narbonolide, narbonolide derivatives and polyketides that
 CC are useful as antibiotics and as intermediates in the synthesis of
 CC compounds with pharmaceutical value. The DNA compounds may also encode
 CC a C12-hydroxylase (pick), desosamine biosynthesis and desosaminyl
 CC transferase enzymes (useful for conversion of ketolides to antibiotics),
 CC and the beta-glucosidase enzyme (involved in picromycin biosynthesis).
 CC These compounds are also useful for increasing the antibiotic activity
 CC of a compound relative to the unhydroxylated compound. The recombinant
 CC host cells are useful as genetic systems that allow rapid engineering
 CC of the narbonolide polyketide synthase. These would be valuable for
 CC creating novel ketolide analogs for pharmaceutical applications.

XX Sequence 1346 AA;

Query Match 6.8%; Score 99.5; DB 21; Length 1346;

Best Local Similarity 20.9%; Pred. No. 0.18;

Matches 67; Conservative 35; Mismatches 107; Indels 111; Gaps 14;

QY 19 EGASLALISPSAVSRQKALEHHVGRVLSRTQPAKATEAG-----EVLVQAR----- 68

Db 688 kggmislalseeatrgrienlh---glslaaavgptatvsgdptqqlaqaceadgir 744

QY 69 -KWVLLQ-----AETRAQLSGRLAEIPLTIAINADSLSTW----- 102

Db 745 ariipdvashshavetieneladvlaspqtpqpfstle----gtwitepalddy 800

QY 103 -----FPVFNVEVASWGAT-----ITLRLDEAHTLSLRRGDLGAVT 142

Db 801 wrynhrvrfapavetlatdegthfievshahpvlmtlpdkvtglatlrred----- 854

QY 143 REANPVAGCEVVELGTMRHLAIATPSLRDAYMYDGLDWAAM-----PVLRFPGKDV 194

Db 855 -----ggqhrll---ttslaeawanglaldwasllpatgalspavdpdiptya 897

QY 195 LQDRDLGDRVDPVGRRRRSIVPSAEGFGEAIRRGLGWGLLPE-----TQAAPML 244

Db 898 fghrsywisppg-geaphtasgreavaet---glawpggaedldeegriravlanvmr 953

QY 245 KAGEVILLD---EIPIDTPM 261

Db 954 gaasvlrcdspeevpdrpl 973

RESULT 10

AAV77195

ID AAY77195 standard; Protein; 1346 AA.

XX AC AAY77195;

XX DT 05-JUN-2000 (first entry)

XX DE S. venezuelae macrolide biosynthetic enzyme PikAIV, SEQ ID NO:37.

XX Desosamine biosynthesis; macrolide; polyketide; methymycin; pikromycin;
 KW neomethymycin; narbomycin; polyhydroxyalkanoate monomer synthase;
 KW biopolymer; antibiotic; chemotherapeutic; immunosuppressant; asthma,
 KW chronic obstructive pulmonary disease; respiratory inflammation;
 KW hypercholesterolaemia; crop protection agent.

OS Streptomyces venezuelae ATCC15439.

XX

PN WO200000620-A2.

XX PD 06-JAN-2000.

XX PF 25-JUN-1999; 99WO-US14398.

XX PR 26-JUN-1998; 98US-0105537.

XX PA (MINU) UNIV MINNESOTA.

XX PI Sherman DH, Liu H, Xue Y, Zhao L;

XX WPI; 2000-160679/14.

XX N-PSDB; AA287300.

XX Desosamine and macrolide biosynthetic gene clusters, useful for, e.g.
 PT synthesis of methymycin and pikromycin
 PS Claim 19; Page 424-428; 438pp; English.

XX The invention relates to an isolated and purified nucleic acid segment
 CC comprising a desosamine biosynthetic gene cluster, a fragment or its
 CC biologically active variant, where the nucleic acid sequence is not
 CC derived from the eryC gene cluster of Saccharopolyspora erythraea or
 CC Streptomyces antibioticus. The invention also relates to a macrolide
 CC biosynthetic gene cluster, or fragments thereof. The macrolide
 CC biosynthetic gene cluster encodes proteins which synthesize methymycin,
 CC pikromycin, neomethymycin, narbomycin or a combination of these
 CC compounds. Recombinant or augmented cells comprising the desosamine
 CC and/or macrolide biosynthetic gene clusters are useful for the
 CC production of biologically active macrolides. The macrolide biosynthetic
 CC proteins are useful for synthesis of methymycin, pikromycin,
 CC neomethymycin and narbomycin. The alternative termination of polyketide
 CC synthesis may be useful to prepare novel antibiotics and
 CC polyhydroxyalkanoate (PHA) monomers. The compounds produced by the
 CC recombinant host cells are useful as biopolymers, e.g., in packaging or
 CC biomedical applications, to engineer PHA monomer synthases or to prepare
 CC biologically active agents, such as chemotherapeutics,
 CC immunosuppressants, agents to treat asthma, chronic obstructive pulmonary
 CC disease as well as other diseases involving respiratory inflammation,
 CC cholesterol-lowering agents or macrolide-based antibiotics which are
 CC active against a variety of organisms, e.g., bacteria, including
 CC multi-drug resistant pneumococci and other respiratory pathogens, as well
 CC as viral parasitic pathogens, or as crop protection agents (e.g.,
 CC fungicides or insecticides) via expression of polyketides in plants.
 CC Sequences AAY77190-Y77197 represent macrolide biosynthetic enzymes from
 CC Streptomyces venezuelae ATCC 15439, which are encoded by sequences
 CC AAZ87295-Z87302.

XX Sequence 1346 AA;

Query Match 6.8%; Score 99.5; DB 21; Length 1346;

Best Local Similarity 20.9%; Pred. No. 0.18;

Matches 67; Conservative 35; Mismatches 107; Indels 111; Gaps 14;

QY 19 EGASLALISPSAVSRQKALEHHVGRVLSRTQPAKATEAG-----EVLVQAR----- 68

Db 688 kggmislalseeatrgrienlh---glslaaavgptatvsgdptqqlaqaceadgir 744

QY 69 -KWVLLQ-----AETRAQLSGRLAEIPLTIAINADSLSTW----- 102

Db 745 ariipdvashshavetieneladvlaspqtpqpfstle----gtwitepalddy 800

QY 103 -----FPVFNVEVASWGAT-----ITLRLDEAHTLSLRRGDLGAVT 142

Db 801 wrynhrvrfapavetlatdegthfievshahpvlmtlpdkvtglatlrred----- 854

QY 143 REANPVAGCEVVELGTMRHLAIATPSLRDAYMYDGLDWAAM-----PVLRFPGKDV 194

Db 855 -----ggqhrll---ttslaeawanglaldwasllpatgalspavdpdiptya 897

QY 195 LQDRDLGDRVDPVGRRRRSIVPSAEGFGEAIRRGLGWGLLPE-----TQAAPML 244

Db 898 fghrsywispagp-geapahtasgreavaet---glawpggaedideegrrsavlamvmr 953

QY 245 KAGEVILLD---EIPIDTPM 261

Db 954 qaasvlrcdspeevpdrpl 973

RESULT 11

ID AAY67204 standard; protein: 1346 AA.

XX AAY67204;

DT 23-MAR-2000 (first entry)

XX Narbonolide synthase subunit 4 (PICAIV) protein sequence.

DE Narbonolide synthase subunit 4 (PICAIV) protein sequence.

XX Narbonolide polyketide synthase; PKS; narbonolide synthase subunit 4;

KW PICAIV; antibiotic production; narbomycin; picromycin; ketolide...

XX Streptomyces venezuelae.

XX WO9961599-A2.

PN 02-DEC-1999.

XX 27-MAY-1999; 99WO-US11814.

XX 28-MAY-1998; 98US-0087080.

PR 28-AUG-1998; 98US-0141908.

PR 22-SEP-1998; 98US-0100880.

PR 08-FEB-1999; 99US-0119139.

XX (KOSA-) KOSAN BIOSCIENCES INC.

XX Ashley G, Betlach MC, Betlach M, McDaniel R, Tang L;

PI WPI; 2000-072618/06.

DR N-PSDB; AAZ56001.

XX New recombinant DNA encoding a domain of narbonolide polyketide synthase, for production of ketolide antibiotics -

PT Example 2; Page 14-15; 98pp; English.

PS This is the Streptomyces venezuelae narbonolide synthase subunit 4, PICAIV protein sequence. The invention relates to recombinant DNA containing a coding sequence for a narbonolide polyketide synthase (PKS). Polyketides are compounds synthesised from 2-carbon units through a series of condensations and subsequent modifications. Modular PKSs are responsible for the production of many antibiotics including picromycin. The narbonolide PKS consists of a loading module, six extender modules, and two thioester domains. Four proteins make up the narbonolide PKS (PICAIV, PICAII, PICAIII and PICAIV). PICAIV includes the loading module and extender modules 1 and 2, PICAII includes extender modules 3 and 4, PICAIII includes extender module 5 and PICAIV includes extender module 6 and a type II thioesterase domain. The second type II thioesterase domain is found on the PICB protein. The nucleotide sequences encoding all of these proteins can be isolated in recombinant form from the recombinant cosmid pKOS023-27 (see AAZ56001). Narbonolide is desosaminylated in S. venezuelae to yield narbomycin, the desosamine transferase enzyme is required for this conversion, and the desosamine biosynthetic genes are also found in cosmid pKOS023-27. The recombinant DNA of the invention is used to express, in transformed cells, narbonolide (or its derivatives) or other ketolides (particularly hybrids), which may then be converted (e.g. by other enzymes recombinantly expressed in the same hosts) to polyketide antibiotics or their intermediates. The antibiotics are useful in human or veterinary medicine.

XX Sequence 1346 AA;

Query Match 6.8%; Score 99.5; DB 21; Length 1346;

Best Local Similarity 20.9%; Pred No. 0.18;

Matches 67; Conservative 35; Mismatches 107; Indels 111; Gaps 14;

QY 19 EGASIALSISPAVSQRVKALEHHVGRVLVSRTQPAKATEAG-----EVLVQAAR----- 68

Db 688 kggmislalseeatqrlienlh---glsiaavngptatvsgdptqigelaqaceadgir 744

QY 69 -KWVLLQ-----AETKAQLSGRLAEIPLTIAINADSLSTW----- 102

Db 745 arlipdyashahvetieneladlagispdpqvpffstle---gtwitepaldggy 800

QY 103 -----PPVFNEVASWGGAT-----LTLRLEDEAHTLSLLRRGDVLAGVT 142

Db 801 wynrlrhrgvfapavetlatdegfthfievshahpvtlmtlpdkvtglatlrred----- 854

QY 143 REANPVAGCEVVELGTMRLHAIATPSLRDAYMVDGKLDWAAM-----PVLRFPGKDV 194

Db 855 -----ggqhrli---ttslaeawangialdwasilpatgalispavdpdlyta 897

QY 195 LODRLDGRVDGPVGRRRVSIYPSAEGFGEAIRRGIGWGLLPE-----TQAAPML 244

Db 898 fghrsywispagp-geapahtasgreavaet---glawpggaedideegrrsavlamvmr 953

QY 245 KAGEVILLD---EIPIDTPM 261

Db 954 qaasvlrcdspeevpdrpl 973

RESULT 12

AAY77180

ID AAY77180 standard; Protein: 12199 AA.

XX AAY77180;

XX 05-JUN-2000 (first entry)

XX S. venezuelae pik gene cluster-encoded protein, SEQ ID NO:6.

XX Desosamine biosynthesis; macrolide; polyketide; methymycin; pikromycin;

KW neomethymycin; narbomycin; polyhydroxyalkanoate monomer synthase;

KW biopolymer; antibiotic; chemotherapeutic; immunosuppressant; asthma,

KW chronic obstructive pulmonary disease; respiratory inflammation;

XX hypercholesterolaemia; crop protection agent.

XX Streptomyces venezuelae ATCC15439.

XX WO200000620-A2.

XX 06-JAN-2000.

XX 25-JUN-1999; 99WO-US14398.

XX 26-JUN-1998; 98US-0105537.

XX (MINU) UNIV MINNESOTA.

XX Sherman DH, Liu H, Xue Y, Zhao L;

XX WPI; 2000-160679/14.

DR N-PSDB; AAZ87285.

XX Desosamine and macrolide biosynthetic gene clusters, useful for, e.g.

PT synthesis of methymycin and pikromycin -

XX Claim 19; Page 315-353; 438pp; English.

XX The invention relates to an isolated and purified nucleic acid segment comprising a desosamine biosynthetic gene cluster, a fragment or its biologically active variant, where the nucleic acid sequence is not derived from the eryC gene cluster of Saccharopolyspora erythraea or Streptomyces antibioticus. The invention also relates to a macrolide

CC biosynthetic gene cluster, or fragments thereof. The macrolide
 CC biosynthetic gene cluster encodes proteins which synthesise methymycin,
 CC pikromycin, neomethymycin, narbomycin or a combination of these
 CC compounds. Recombinant or augmented cells comprising the desosamine
 CC and/or macrolide biosynthetic gene clusters are useful for the production
 CC of biologically active macrolides. The macrolide biosynthetic proteins
 CC are useful for synthesis of methymycin, pikromycin, neomethymycin and
 CC narbomycin. The alternative termination of polyketide synthesis may be
 CC useful to prepare novel antibiotics and polyhydroxyalkanoate (PHA)
 CC monomers. The compounds produced by the recombinant host cells are useful
 CC as biopolymers, e.g., in packaging or biomedical applications, to
 CC engineer PHA monomer synthases or to prepare biologically active agents,
 CC such as chemotherapeutics, immunosuppressants, agents to treat asthma,
 CC chronic obstructive pulmonary disease as well as other diseases involving
 CC respiratory inflammation, cholesterol-lowering agents or macrolide-based
 CC antibiotics which are active against a variety of organisms, e.g.,
 CC bacteria, including multi-drug resistant pneumococci and other
 CC respiratory pathogens, as well as viral parasitic pathogens, or as crop
 CC protection agents (e.g., fungicides or insecticides) via expression of
 CC polyketides in plants. The present sequence represents a protein
 CC encoded by the macrolide biosynthetic gene cluster (pik) from
 CC Streptomyces venezuelae ATCC 15439.
 XX
 SQ Sequence 12199 AA;

Query Match 6.8%; Score 99.5; DB 21; Length 12199;
 Best Local Similarity 20.9%; Pred. No. 4.6;
 Matches 67; Conservative 35; Mismatches 107; Indels 111; Gaps 14;

QY 19 EGASLALISPSAVSQRKALEHHVGRVLSRTPAKATEAG-----EVLVQAR----- 68
 Db 11260 kgmislalseeatrgrienh---glisiaavngptatvsgdptqigelaqaeadgir 11316
 QY 69 -KVVLLQ-----AETKAOLSGRLAEIPLTIAINADSLSTW----- 102
 Db 11317 ariipdvashshvetieneladvlaglspqtpqvffstle---gtwitepalddgy 11372
 QY 103 -----FPPVFNESVWGKAT-----LFLRLEDEAHTLSLLRRGDLGAVT 142
 Db 11373 wrynrlhrvgfapavetlatdegthfievshapvlmtipdkvtglatirred----- 11426
 QY 143 REANPVAGCEVVELGTRHLAIATPSLRDAYWVDGKLDWAM-----PVLRFPGKDV 194
 Db 11427 -----ggqhr1---ttsiaewanglaldwasllpatgalspavpdlptya 11469
 QY 195 LQDRDLGRVDGPGVRRRSVTSVSAEGFGAIRRGLGWGLLPE-----TQAAPML 244
 Db 11470 fhrsywispagp-geapahtasgreavaet---glawggaedldegrrsavlanvmr 11525
 QY 245 KAGEVILLD---EIPIDTPM 261
 Db 11526 qaasvlrcdspeevpvdrl 11545

RESULT 13

AAW74591
 ID AAW74591 standard; Protein; 303 AA.

XX
 AC AAW74591;

XX
 DT 21-MAR-2000 (first entry)

XX
 DE Neisseria gonorrhoeae ORF 158 protein sequence SEQ ID NO:656.

XX
 KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
 KW antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
 KW antibacterial; gene therapy.

XX
 OS Neisseria gonorrhoeae.

XX
 PN WO957280-A2.

XX

PD 11-NOV-1999.

XX
 PF 30-APR-1999; 99WO-US09346.

XX
 PR 01-MAY-1998; 98US-0083758.

XX
 PR 31-JUL-1998; 98US-0094869.

XX
 PR 02-SEP-1998; 98US-0098994.

XX
 PR 09-OCT-1998; 98US-0103749.

XX
 PR 09-OCT-1998; 98US-0103794.

XX
 PR 09-OCT-1998; 98US-0103796.

XX
 PR 25-FEB-1999; 99US-0121528.

XX
 PA (CHIR) CHIRON CORP.

XX
 PA (GENO-) INST GENOMIC RES.

XX
 PI Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;

XX
 PI Petersen J, Pizza M, Rappuoli R, Ratti G, Scalato E, Scarselli M;

XX
 PI Tettelin H, Venter JC;

XX
 DR WPI; 2000-062150/05.

XX
 DR N-PSDB; AAZ53353.

XX
 PT Novel Neisserial polypeptides predicted to be useful antigens for

XX
 PT vaccines and diagnostics

XX
 PS Claim 2; Page 452; 1453pp; English.

XX
 CC AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941

XX
 CC represent novel Neisseria meningitis and N. gonorrhoeae polynucleotides

XX
 CC and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ55473 represent

XX
 CC PCR primers used in the exemplification of the present invention. The

XX
 CC polypeptides, the polynucleotides, antibodies and compositions of

XX
 CC the invention can be used as vaccines, as diagnostic reagents, and as

XX
 CC immunogenic compositions. The polypeptides can be used in the

XX
 CC manufacture of medicaments for treating or preventing infection due to

XX
 CC Neisserial bacteria (e.g. meningitis and septicaemia), to detect the

XX
 CC presence of Neisseria bacteria, or to raise antibodies. They may also

XX
 CC be used to screen for agonists or antagonists, which may themselves

XX
 CC have use as antibacterial agents. The polynucleotides of the invention

XX
 CC may also be used in gene therapy protocols.

XX
 SQ Sequence 303 AA;

Query Match 6.7%; Score 98; DB 21; Length 303;
 Best Local Similarity 27.6%; Pred. No. 0.029;
 Matches 43; Conservative 28; Mismatches 55; Indels 30; Gaps 7;

QY 2 NPQLDTLLSIIDEGSFEGLASLALISPSAVSQRKALEHHVGRVLSR-TQPAKATEAG 60

Db 4 nseeltvfvqvesgsfrraaeqlemansavsrivkrleeklgvnlrrtrglnlteeg 63

QY 61 EVLVQARKKVVLLQAEETKAOLSGRLAEIPLTIAINADSLSTWPPV-----FNE- 109

Db 64 aqyfriraqrlqemaetaetema-vhevpggv-irvda---mpmvhlhllaplaakfner 118

QY 110 -----VASWGGATLTLRLEDEAHTLSLLRRGDV 137

Db 119 yphirlslvssegynlierkvdia-----lragel 149

RESULT 14

AAW52846

ID AAW52846 standard; Protein; 5069 AA.

XX
 AC AAW52846;

XX
 DT 24-JUL-1998 (first entry)

XX
 DE A. mediterranei rifamycin synthesis gene cluster fragment protein B.

XX
 KW Amycolatopsis mediterranei; rifamycin; synthesis; gene cluster;

KW polyketide synthase; actinomycete; ansamycin.

OS Amycolatopsis mediterranei.

PN W09807868-Al.

XX 26-FEB-1998.

XX 18-AUG-1997; 97WO-EP04495.

XX 20-AUG-1996; 96EP-0810551.

XX (NOVS) NOVARTIS AG.

XX Engel N, Schupp T, Toupet C;

XX WPI: 1998-169172/15.

XX DR N-PSDB; AAV211187.

XX Amycolatopsis mediterranei rifamycin synthesis gene cluster - used
PT to produce rifamycin and rifamycin analogues

XX Claim 6; Page 126-151; 205pp; English.

XX The present sequence represents a Amycolatopsis mediterranei rifamycin
CC synthesis gene cluster ORF B protein from the present invention. The
CC DNA fragment comprises a DNA region involved directly or indirectly
CC in the gene cluster responsible for rifamycin synthesis, including
CC the adjacent DNA regions to the right and left which, by reason of
CC their function in connection with rifamycin biosynthesis, qualify
CC as constituents of this rifamycin gene cluster, and functional
CC fragments, derivatives or constituents of these. The Amycolatopsis
CC mediterranei rifamycin synthesis gene cluster DNA fragment can be used
CC for producing rifamycin, rifamycin analogues or precursors. It can also
CC be used for inactivating or modifying genes involved in ansamycin or
CC rifamycin biosynthesis. The DNA can be used for constructing mutant
CC actinomycetes strains from which the natural rifamycin or ansamycin
CC biosynthesis gene cluster has been partly or completely deleted. The
CC DNA fragment can be used for assembling a library of polyketide
CC synthases, which can be used for assembling a library of polyketides.
CC A hybridisation probe of the invention can be used for identifying DNA
CC fragments involved in the biosynthesis of ansamycins.

XX Sequence 5069 AA;

Query Match 6.5%; Score 95.5; DB 19; Length 5069;
Best Local Similarity 23.7%; Pred. No. 3.4; Mismatches 79; Gaps 15;
Matches 71; Conservative 45;

QY 21 ASIALSISPSAVSQRV-----KALEHHVGRVLSVRTQPAKATEAGEVLVQAAARKWVLLQA 75

Db 3919 asfamvvglaavqsgvgrpdavghsqgei-----aaacvsgalslqdaakvvalrs 3971

QY 76 EYKA-QLSGR--LAELPLTIANADSLSTWTFPP-----VFNEVASWGGATLRLRDE 124

Db 3972 qalatrlagrgma-----svalseedatawlapwdrqvaaavnspsavviageaqalde 4027

QY 125 EAHTLS---LLRRGDV-LGAVTRE-----ANPVACEVVELGTMRLHAIATPSLRD 171

Db 4028 vvdalsgqevrvavdygshntqvealedllaetlagieaqkvpfyscllgdwird 4087

QY 172 AYMDGKLDWAAMPVLRFGPKDVLQDRDLGRDGVGVRGRRRSIVPSAEGFEAIRRGL- 230

Db 4088 agivdgyggy-wy-----rnlrnqv-----gfgpavaelvr 4115

QY 231 -CWGLLPETQAAPMLKAGEVILLDEIPDTPMWNORWRLSRLARLTDAVDAIEGL 288

Db 4116 qdngvfvevsahpvl-----vqplselis-ddavvtgslrrredgglrrlitsmaelyvqgv 4169

RESULT 15
c AAR25588

ID AAR25588 standard; Protein; 298 AA.

XX AAR25588;

XX 08-JAN-1993 (first entry)

XX Truncated irgB.

XX Virulence factor; intestine; multi-valent vaccine.

XX Vibrio cholerae.

XX W09211354-A.

XX 09-JUL-1992.

XX 18-DEC-1991; 91WO-US09592.

XX 18-DEC-1990; 90US-0629102.

XX (GEHO) GEN HOSPITAL CORP.

XX (HARD) HARVARD COLLEGE.

XX Calderwood SB, Goldberg MB, Mekalanos JJ;

XX WPI: 1992-250077/30.

XX DR N-PSDB; AAQ26543.

XX Cholera vaccine comprising live attenuated Vibrio cholerae cells
PT - contains mutations which inhibit expression of a functional
PT irgA gene product, vaccine has reduced virulence, decreasing side
PT effects

XX Disclosure; Fig 13; 75pp; English.

XX The functional irgB gene encodes a protein which acts as a positive
CC regulator of irgA. irgA functions as a V. cholerae virulence factor.
CC The sequence given shows a truncated irgB. This truncation of the
CC sequence inhibits expression of a functional irgA gene product in the
CC cell. This sequence can be used to transform V. cholerae cells which
CC can then be used to induce immunity to cholera in a mammal. These
CC transformed cells can colonize the intestines of inoculated animals and
CC may be combined with vaccines targeted at other illnesses to make a
CC single multi-valent vaccine. The transformed cells have reduced
CC virulence such that the inoculated animal is less likely to develop
CC side effects such as diarrhoea and fever.

XX Sequence 298 AA;

Query Match 6.4%; Score 93.5; DB 13; Length 298;

Best Local Similarity 23.0%; Pred. No. 0.087; Mismatches 33; Indels 71; Gaps 11;
Matches 59; Conservative 33;

QY 17 SFEGLASLISPSAVSQRVKALEHHVGRVLSVRT-OPAKATEAGEV-----LVQAAAR 68

Db 18 sitaaakaleqkstklsrrlaqleedlqgslImrggnriltkagevfayvseqillelan 77

QY 69 KVVLLQAEKTKAQLSGLRLEIPLTIANADSLSTWTFPPVFNEVASWGGATLRL----- 122

Db 78 Ksqaalqelnngvtge-----ltlvvhpnlirgwlsvqvidefmq-ghstcklrllsqfgh 131

QY 123 EDEAHTLSLLRRGDVILGAVTREANPNVAGCEVVELGTMRLHAIATPSL-----RDAYMVDGKL 179

Db 132 sdevlepdli-----iwichaapmgyrkerlygyryatyaspkylahrd----- 175

QY 180 DWAAAMPVLRFGPKDVLQDRDLGRDGVGVRGRRRSIVPSAEGFG----- 223

Db 176 ----kpt---hprelihhpwidf-----iacrraelelhphfsgyslpalesrlqsdnl 223

QY 224 ----EAIIRGLGWGLLP 236

Db 224 amqadaiaakgrgigllp 240

Search completed: September 13, 2001, 17:51:47
Job time: 1118 sec

Result No.	Score	Query		Length	DB	ID	Description
		Match	%				
1	136.5	9.3	276	1	US-07-952-817-27	Sequence 27, Appl	
2	118.5	8.1	278	4	US-09-036-987A-22	Sequence 22, Appl	
3	99.5	6.8	1346	3	US-09-320-878-4	Sequence 4, Appl	
4	96	6.6	251	1	US-07-952-817-28	Sequence 28, Appl	
5	92.5	6.3	4150	4	US-09-428-517-2	Sequence 2, Appl	
6	89	6.1	290	2	US-08-614-686A-1	Sequence 1, Appl	
7	89	6.1	840	2	US-08-467-822-25	Sequence 25, Appl	
8	89	6.1	840	4	US-08-432-697-25	Sequence 25, Appl	
9	87.5	6.0	433	4	US-09-385-028-10	Sequence 10, Appl	
10	87.5	6.0	3798	3	US-09-335-409-6	Sequence 6, Appl	
11	86.5	5.9	3031	1	US-07-689-008-2	Sequence 2, Appl	
12	85.5	5.9	291	1	US-07-952-817-26	Sequence 26, Appl	
13	85	5.8	1068	1	US-08-537-210A-2	Sequence 2, Appl	
14	85	5.8	1068	4	US-09-113-825-2	Sequence 2, Appl	
15	85	5.8	2556	1	US-08-185-432-17	Sequence 17, Appl	
16	84.5	5.8	3739	3	US-09-320-878-2	Sequence 2, Appl	
17	84	5.8	829	4	US-09-413-814-105	Sequence 105, App	
18	84	5.8	1548	1	US-08-463-092B-7	Sequence 7, Appl	
19	84	5.8	1548	2	US-08-460-907B-7	Sequence 7, Appl	
20	83.5	5.7	591	4	US-09-413-814-76	Sequence 76, Appl	
21	83.5	5.7	3816	4	US-09-428-517-3	Sequence 3, Appl	
22	82.5	5.7	2152	4	US-09-036-987A-3	Sequence 3, Appl	
23	82.5	5.7	4472	2	US-08-804-227C-2	Sequence 2, Appl	
24	81.5	5.6	503	3	US-08-911-853-7	Sequence 7, Appl	
25	81.5	5.6	503	4	US-09-479-409-7	Sequence 7, Appl	
26	81.5	5.6	2509	1	US-08-469-005A-10	Sequence 10, Appl	
27	81.5	5.6	4545	2	US-08-804-227C-14	Sequence 14, Appl	


```

CITY: Roslyn
STATE: New York
COUNTRY: U.S.A.
ZIP: 11576
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WordPerfect Version 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/614,686A
FILING DATE: MARCH 12, 1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GR195110930
FILING DATE: 24 MARCH 1995
APPLICATION NUMBER: GR195114056.7
FILING DATE: 13 APRIL 1995
ATTORNEY/AGENT INFORMATION:
NAME: Collard, Allison C.
REGISTRATION NUMBER: 22,532
REFERENCE/DOCKET NUMBER: BOCK ET AL. -4
ATTORNEY/AGENT INFORMATION:
NAME: Freedman, Edward R.
REGISTRATION NUMBER: 26,048
REFERENCE/DOCKET NUMBER: BOCK ET AL. -4
ATTORNEY/AGENT INFORMATION:
NAME: Keusey, Edwin H.
REGISTRATION NUMBER: 34,361
REFERENCE/DOCKET NUMBER: BOCK ET AL. -4
TELECOMMUNICATION INFORMATION:
TELEPHONE: (516) 365-9802
TELEFAX: (516) 365-9805
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 290 amino acids
TYPE: amino acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Klebsiella terrigena
STRAIN: DSM2867
US-08-614-686A-1

```

Query Match 6.1%; Score 89; DB 2; Length 290;
 Best Local Similarity 23.6%; Pred. No. 0.071;
 Matches 74; Conservative 56; Mismatches 105; Indels 78; Gaps 20;

```

QY 10 LSTIDSGFEGASLALSISPSAVSORVKALEHHVGRVLSR-TOPAKATEAGE-VLVQAA 67
Db 10 VAVAEARNTRAHDIGISQPLSQIQRLEIGTGLRLTRGVLTGAGESFYVDAC 69
QY 68 RMVLLOA---ETKA---OLSGRLAE-IPLTAINADSLSTW-----PPPVFNEVASWG 114
Db 70 QIALSDAALEKTKGIARGMWSLVPITSSAAFSHQIESLLYQFOQRYPAVALRQVEGN 129
QY 115 GATLT-----LRL---EDEAHTLSLLRCDVLGAVTREPANPVAGCEVVELGTM 159
Db 130 MATLHMALGEAEIDTAFTAPCESSKAFNLRIIAEPMVIALHR-SHPLUG----- 179
QY 160 RHLATATPSLRDAYMVDGKLDWAAMPVLRFGPKDV--LQDRDLG--RVDPGVGR-RRVS 214
Db 180 -ESALSQAQSD-----AVPVI-FPEVAPGLYEQVYDGCRRAGVDMSSRQSS 226
QY 215 IYPSAGGFEATRGWGGLLPETQAA-----PMLKAGEVILLDDIPIDTMYWQR 265
Db 227 QISSISIMVDA---GFGFALVPQSMTCICLPNVWTHPLQDAS---LKTETAI-----AWRR 276
QY 266 WRLESRLARLTD 278
Db 277 FE-RSETVKRFL 288

```

RESULT 7

```

US-08-467-822-25
; Sequence 25, Application US/08467822
; Patent No. 5843460
; GENERAL INFORMATION:
; APPLICANT: Labigne, Agnes
; APPLICANT: Sauerbaum, Sebastien
; APPLICANT: Ferrero, Richard L.
; APPLICANT: Thiberge, Jean-Michel
; TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS AGAINST
; TITLE OF INVENTION: HELICOBACTER INFECTION, POLYPEPTIDES FOR USE IN THE
; TITLE OF INVENTION: COMPOSITIONS, AND NUCLEIC ACID SEQUENCES ENCODING SAID
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
; ADDRESSEE: Dunner
; STREET: 1300 I Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/467,822
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/447,177
; FILING DATE: 19-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/432,697
; FILING DATE: 02-MAY-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Meyers, Kenneth J.
; REGISTRATION NUMBER: 25,146
; REFERENCE/DOCKET NUMBER: 03495.0137-02000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 408-4000
; TELEFAX: (202) 408-4400
; INFORMATION FOR SEQ ID NO: 25:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 840 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-467-822-25

```

Query Match 6.1%; Score 89; DB 2; Length 840;
 Best Local Similarity 22.1%; Pred. No. 0.36;
 Matches 59; Conservative 30; Mismatches 114; Indels 64; Gaps 10;

```

QY 34 QRVKALEHHVGRVLSRTPAKATEAGEVLVQAAKRWLLQAEKQALSGRLAEIPLTIA 93
Db 69 QVLPVAPHLNVAQVETEP-----DGTKLVTVDHPISRENGELQALFGLSLPVP--- 119
QY 94 INADSLSTWFPVFNVEVASWGATLTLRLEDEAHTLSLLRGRDVLGAVTREPANPVAGCV 153
Db 120 ----SLDK-----FAETKEDNRPGEILCEDECLTINIGRKAVILKVTSGKDRP----- 164
QY 154 VELGTMRLHAIATPSL-----RDAYMVDGKLDWAAMPVLRFGPKDVLQDR--DLQGR--- 203
Db 165 IOVGSYHFIEVNPVLTDFDRRKAYGM--RLNIAAGTAVRFEFGDCKSVTLVSIEGNKVR 222

```

QY 204 -----VDGPVRRRSTVPSA---EGFGEAIRRLGWLGLLPE----- 237
Db 223 GGNAIADGPNVNTLEAMHVRSGFGHEEKDAPEGFTTKEDPNCFSNTFIHRKEYANK 282
QY 238 --TQAAPMLKAGEVILLDEIPIDTPMY 262
Db 283 YGPTTGDKIRLIGDNTLLAEIEKDYALY 309

RESULT 8
US-08-432-697-25
; Sequence 25, Application US/08432697
; Patent No. 6248330
; GENERAL INFORMATION:
; APPLICANT: Labigne, Agnes
; APPLICANT: Sauerbaum, Sebastien
; APPLICANT: Ferrero, Richard L.
; APPLICANT: Thiberge, Jean-Michel
; TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS AGAINST
; TITLE OF INVENTION: HELICOBACTER INFECTION, POLYPEPTIDES FOR USE IN THE
; TITLE OF INVENTION: COMPOSITIONS, AND NUCLEIC ACID SEQUENCES ENCODING SAID
; TITLE OF INVENTION: POLYPEPTIDES
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
; STREET: 1300 I Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/432,697
; FILING DATE: 02-MAY-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Meyers, Kenneth J.
; REGISTRATION NUMBER: 25,146
; REFERENCE/DOCKET NUMBER: 03495.0137-00000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 408-4000
; TELEFAX: (202) 408-4400
; INFORMATION FOR SEQ ID NO: 25:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 840 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-432-697-25

Query Match 6.1%; Score 89; DB 4; Length 840;
Best Local Similarity 22.1%; Pred. No. 0.36;
Matches 59; Conservative 30; Mismatches 114; Indels 64; Gaps 10;

QY 34 QRVKALEHHVGRVILVSRTPAKATEAGEVLVQAARKMVLQLQAEAKQALSGRLAIEPLTIA 93
Db 69 QVLPVPHLLNAVQVEATEP-----DGTCLVTHDPTISRENGELQAEALFGLSLPVP----- 119
QY 94 INADSLSTWPPVNFVNASWGAGTATLURLDEAHTLSLLRRGDVLGAVTREANPVACEY 153
Db 120 ----SLDK-----FAETKEDNRPGEILCEDECLTLNIGRKAVILKVLTSKGRDP----- 164
QY 154 VELGTMRHLAIAIATPSL-----RDAYWVDGKLDWAAMPVLRFGPKDVLQDR--DLDDR----- 203
Db 165 IQVGSYHFIEVNPYLTDFDRRKAYGM--RLNIAAGTAVRFPDGDCKSVTLVLSIEGKVKIR 222

QY 204 -----VDGPVRRRSTVPSA---EGFGEAIRRLGWLGLLPE----- 237
Db 223 GGNAIADGPNVNTLEAMHVRSGFGHEEKDAPEGFTTKEDPNCFSNTFIHRKEYANK 282
QY 238 --TQAAPMLKAGEVILLDEIPIDTPMY 262
Db 283 YGPTTGDKIRLIGDNTLLAEIEKDYALY 309

RESULT 9
US-09-385-028-10
; Sequence 10, Application US/09385028
; Patent No. 6232106
; GENERAL INFORMATION:
; APPLICANT: Susan E. Jensen
; APPLICANT: Kwamena A Aidoo
; APPLICANT: Ashish S. Paradkar
; TITLE OF INVENTION: DNA Sequence Encoding Enzymes of Clavulanic
; Patent No. 6232106
; TITLE OF INVENTION: Acid Biosynthesis
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: JACOBSON, PRICE, HOLMAN & STERN, PLLC
; STREET: The Jennifer Bulding, 400 Seventh Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30 (BPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/385,028
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA: US 08/790,462
; APPLICATION NUMBER: US 08/790,462
; FILING DATE: 29-JAN-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: D. Douglas Price
; REGISTRATION NUMBER: 24,514
; REFERENCE/DOCKET NUMBER: 1418/P57452US2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 638-6666
; TELEFAX: (202) 39305350
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 433 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-385-028-10

Query Match 6.0%; Score 87.5; DB 4; Length 433;
Best Local Similarity 33.7%; Pred. No. 0.19;
Matches 29; Conservative 16; Mismatches 38; Indels 3; Gaps 2;

QY 6 LDTLLSIDSGSEGASLALSISPSAVSQVRKALEHHVGRVILVSRTPAKATEAGEVLVQ 65
Db 330 LHLQAVARHGSINRAAVLISQSALTRHRLQESIGARLLRS--PRGTSLTGTRQ 387
QY 66 AARKMVLQAE--TKAQLSGRLAIEPL 90
Db 388 FLRLQALYAEFFREAAALACRSVERPL 413

RESULT 10

US-09-335-409-6

Sequence 6, Application US/09335409

Patent No. 6121029

GENERAL INFORMATION:

APPLICANT: Schupp, Thomas

APPLICANT: Ligon, James

APPLICANT: Molnar, Istvan

APPLICANT: Zirkle, Ross

APPLICANT: Cyr, Devon

APPLICANT: Goerlach, Joern

TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES

FILE REFERENCE: 4-30582A

CURRENT APPLICATION NUMBER: US/09/335,409

CURRENT FILING DATE: 1999-06-17

NUMBER OF SEQ ID NOS: 30

SOFTWARE: Patent In Ver. 2.0

SEQ ID NO 6

LENGTH: 3798

TYPE: PRT

ORGANISM: Sorangium cellulosum

US-09-335-409-6

Query Match

Best Local Similarity 6.0%; Score 87.5; DB 3; Length 3798;

Matches 82; Conservative 29; Mismatches 117; Indels 141; Gaps 16;

QY 2 NPQLDTLLSIIDEGSFEGLASLSISPAVSQVKALE-----HH---42

Db 345 DPEIEALRAVVGPARADGRCVL-----GAVKTNLHLEGAAGVAGLIKATLSLHHERIP 400

QY 43 -----VGRVLVSRTPA-----KATEAGEVLVQAARKM 70

Db 401 RNLNFTLNPRIEGTALALATEPVWPTGRTREAGVSGMGTNAHVLEEAP---457

QY 71 VLLQATKQALSGRLAIEPLTIAINADSLSTWPPVNEVASWGGATLRLLEDEAHTLS 130

Db 458 ---AVEPEAAAPRAAELEFVLSAKSAALD-----QAARLRDHLKXHV 498

QY 131 LLRRGQVLG--AVTREANVAGCEVVELGTMRH-LAIATPS---LRDAY-----173

Db 499 ELGLDVAFLATRSA-----MEHRLVAASSREALGALSAAGHTTPG 545

QY 174 MYDGKLDWAAMPVLR-----GPKDVLQDRDLGRVDPVGRRRVSIVPSAEGFGEAIRR 228

Db 546 AVGRASGSGAPKVVFPFGQGSOWGMGRKL--MAEPPVFR-----AALEGCDRAIEA 597

QY 229 GLCWGLLPETQA-----APMLKAGEVILLDEIPDTPMYQWRWRLSESLAR 275

Db 598 EAGWSLLGELSADAASQLGRIDVQPVLFAMEVAL-----SALWRSWGEPEAVVG 649

QY 276 LTDVAVYDAA 284

Db 650 HSMGEVAAA 658

RESULT 11

US-07-689-008-2

Sequence 2, Application US/07689008

Patent No. 5268274

GENERAL INFORMATION:

APPLICANT: Ben-Bassat, Arie

APPLICANT: Calhoon, Roger D

APPLICANT: Fear, Anna L

APPLICANT: Gelfand, David H

APPLICANT: Meade, James H

APPLICANT: Tal, Ronny

APPLICANT: Wong, Hing

APPLICANT: Benzman, Moshe

TITLE OF INVENTION: METHODS AND NUCLEIC ACID SEQUENCES FOR THE

TITLE OF INVENTION: EXPRESSION OF CELLULOSE SYNTHASE OPERON

NUMBER OF SEQUENCES: 15

CORRESPONDENCE ADDRESS:

ADDRESSEE: McCutchen, Doyle, Brown & Enersen

STREET: Three Embarcadero Center

CITY: San Francisco

STATE: California

COUNTRY: USA

ZIP: 94111

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.24

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/689,008

FILING DATE: 19910422

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 337,194

FILING DATE: 12-APR-1989

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 496,236

FILING DATE: 23-MAR-1990

ATTORNEY/AGENT INFORMATION:

NAME: Murphy, Lisabeth Feix

REGISTRATION NUMBER: 31547

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 393-2000

TELEFAX: (415) 393-2286

TELEX: 340817 MACPAG SFO

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 3031 amino acids

TYPE: AMINO ACID

TOPOLOGY: linear

MOLECULE TYPE: protein

US-07-689-008-2

Query Match

Best Local Similarity 5.9%; Score 86.5; DB 1; Length 3031;

Matches 68; Conservative 36; Mismatches 94; Indels 117; Gaps 14;

QY 23 LALSTSPSAVSQVRKAL--EHHVGRVLSR-----TOPAKATEAGEVLVQ 65

Db 2120 LLAGLSPADYSPAIRSIAEMEIKDLSRLSMVSNPVLIREALTQPDPTGARGVAVAD 2179

QY 66 AAR-----KWLLOAETKQALSGRLAEIPLTIAINAD---SLSTWPPVNEVAS--112

Db 2180 LFRQRGDMVHARMALRIASR-----TIDLSPQRLSYATEYMKISNPVAAAR 2227

QY 113 -----WGGATLTLRLEDEAHTLSLLRRG-----DVLG-----AVTREAN 146

Db 2228 LLAPLGDGTSATGSALLPEQVTLQQLRMGISVAGSDLLNQRGDQAQAYDHLAPALQAD 2287

QY 147 PVAGCEVVELGTMRHLAATPSLRDAYWVDGKLDWAMPVLRFGPKD-----VL 195

Db 2288 PEA-----TSPKLALA--RLYNGHKGPKGKALETDLAVLRHNPQDLDARQAQVAAVYN 2337

QY 196 QDRD-----LDGRVDGPVGR--RVSVIPSAEGFG-----EAIR 227

Db 2338 SDHNSLATFLAMDGVQESPMADARAWLAMAVADQADGHGORTIEDLRLRDLRLOQVEGTR 2397

QY 228 RGLGWGLLPETQAAP 242

Db 2398 AASGAGAAQEDALAP 2412

RESULT 12

US-07-952-817-26

Sequence 26, Application US/07952817

Patent No. 5356796

GENERAL INFORMATION:

APPLICANT: Keller, John W.

TITLE OF INVENTION: A Repressor Protein and Gene for Regulating

;; TITLE OF INVENTION: Expression of Polypeptides and Its Use in the Preparation of
;; TITLE OF INVENTION: 2,2-Dialkylglycine Decarboxylase of Pseudomonas Cepacia
;; NUMBER OF SEQUENCES: 30
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
;; ADDRESSEE: Dunner
;; STREET: 1300 I Street, N.W., Suite 700
;; CITY: Washington
;; STATE: D.C.
;; COUNTRY: US
;; ZIP: 20005-3315
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patent In Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/07/952.817
;; FILING DATE: 19920928
;; CLASSIFICATION: 435
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Meyers, Kenneth J.
;; REGISTRATION NUMBER: 25,146
;; REFERENCE/DOCKET NUMBER: 01120.0002-01000
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 202-408-4000
;; TELEFAX: 202-408-4400
;; INFORMATION FOR SEQ ID NO: 26:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 291 amino acids
;; TYPE: AMINO ACID
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
;; US-07-952-817-26

Query Match 5.9%; Score 85.5; DB 1; Length 291;

Best Local Similarity 21.8%; Pred. No. 0.18;
Matches 63; Conservative 44; Mismatches 113; Indels 69; Gaps 11;

Qy 1 MNPQLDLSIIIDEGSFGASLALSPSAVSORVKALEHHVGRVLVSRTPAKATEA 59
Db 1 MNIKDLEVLALAEHRFRRAADSCHVSQPTLSGQIRKLEDELGWMLETSRVLFQA 60
Qy 60 GEVLVQARKMVLQAETKAQLSGLAEI---PLTIAINADSLSTWFPVNEVASWGGA 116
Db 61 GMLLVQAR-TVREVKVLKEMASQOGETMSGPLHGL---IPTVGYLLPHIIPMLHQ 115
Qy 117 TLTRLE---DEAHTLSLLRGD-----VLGAVTREA-----NP 147
Db 116 TFP-KLEMYLHEAQTQALLAQLDCKLDCVTLALVKESERFIEVPLFDEPMLAIYEDHP 174
Qy 148 VAGCEVVELGTMRLHAIATPSLRDAYVWDGKLDWAAMPVLRFGPKDVLDRDL----DGR 203
Db 175 WANRECVPM-----DIAGEKLLMLDGHCLRDQAMGCFEAG 212
Qy 204 VDGPGVRRRVSVPSABGFGBAIRGLGWGLLPETQAAPMLKAGEVILL 252
Db 213 ADEDTFRATSL---ETLRNVAAGSGITLLPALAVPPERKRGVYL 257

RESULT 13

US-08-537-210A-2
; Sequence 2, Application US/08537210A
; Patent No. 5780300
; GENERAL INFORMATION:
; APPLICANT: Artavanis-Tsakonas, Spyridon
; APPLICANT: Fortini, Mark
; APPLICANT: Matsuno, Kenji
; TITLE OF INVENTION: MANIPULATION OF NON-TERMINALLY
; TITLE OF INVENTION: DIFFERENTIATED CELLS USING THE NOTCH PATHWAY
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:

;; ADDRESSEE: Pennie & Edmonds
;; STREET: 1155 Avenue of the Americas
;; CITY: New York
;; STATE: NY
;; COUNTRY: USA
;; ZIP: 10036/2711
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Diskette
;; COMPUTER: IBM Compatible
;; OPERATING SYSTEM: DOS
;; SOFTWARE: FastSeq Version 2.0
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/537,210A
;; FILING DATE: 29-SEP-1995
;; CLASSIFICATION: 435
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Mirock, S. Leslie
;; REGISTRATION NUMBER: 18,872
;; REFERENCE/DOCKET NUMBER: 7326-027
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 212-790-9090
;; TELEFAX: 212-869-8864
;; TELEX: 66141 PENNIE
;; INFORMATION FOR SEQ ID NO: 2:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 1068 amino acids
;; TYPE: amino acid
;; STRANDEDNESS:
;; TOPOLOGY: unknown
;; MOLECULE TYPE: protein
;; FEATURE:
;; NAME/KEY: Human N1 (TAN-1)
;; LOCATION: 1152...2219
;; OTHER INFORMATION: Highly conserved ankryrin repeat
;; OTHER INFORMATION: region of No. 5780300ch
;; US-08-537-210A-2

Query Match 5.8%; Score 85; DB 1; Length 1068;

Best Local Similarity 21.6%; Pred. No. 1.5;
Matches 70; Conservative 44; Mismatches 106; Indels 104; Gaps 16;

Qy 11 SIIDEGSFGASLALSPSAVSORVKALEHHVGRVLVSRTPAKATEAGEVLVQAARKM 70
Db 758 AVISDFIYQASL-----HNQDRTGTETALHLAAR--YSRSDAAKR----- 796
Qy 71 VLLQAEYKAQLSGLAEIPLTIAINADSLSTWFPVNEVASWGAT-LTLRLEDEAHTL 129
Db 797 -LLEASADANTQDNMGRTPLHAAVSADAQGVF-----QILINRATDLDARMHDGTTPL 849
Qy 130 SLLRRGDVLGAVTREANPVAGCEVV-ELG-TMRHLATATPSLRDAYV----- 175
Db 850 ILAARLAVEGMLEDLINSADVNAVDDLGKLSALHWAANNV-DAVVLLKNGANKDMQN 908
Qy 176 -----DGKLDWAAMPVLRFG-----PKDVLODR---DLGGRVD----- 205
Db 909 NREETPLFLAAWEGSYETAKVLLDHFANWDITDHMDRLPRDIAQERMHHDIVRLLEYNL 968
Qy 206 -----GPGVRRRVSVIP--SAEGF-----GEATRRGLGWGLLPETQAAPMLKA 246
Db 969 VRSPQLHGAPLGGTPTLSPPLCSPNGYLSLKPVGQKKVRKPSKGLACGSKAKDLKA 1028
Qy 247 -----GEVILLDE---IPIDT 259
Db 1029 WRKKSQDGKGLDSSGMLSPVDS 1052

RESULT 14

US-09-113-825-2
; Sequence 2, Application US/09113825
; Patent No. 6149902
; GENERAL INFORMATION:
; APPLICANT: Artavanis-Tsakonas, Spyridon

APPLICANT: Fortini, Mark
 APPLICANT: Matsuno, Kenji
 TITLE OF INVENTION: MANIPULATION OF NON-TERMINALLY
 TITLE OF INVENTION: DIFFERENTIATED CELLS USING THE NOTCH PATHWAY
 NUMBER OF SEQUENCES: 4
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Pennie & Edmonds
 STREET: 1155 Avenue of the Americas
 CITY: New York
 STATE: NY
 COUNTRY: USA
 ZIP: 10036/2711
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette
 COMPUTER: IBM Compatible
 OPERATING SYSTEM: DOS
 SOFTWARE: FastSeq Version 2.0
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/113,825
 FILING DATE:
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/537,210
 FILING DATE: 29-SEP-1995
 ATTORNEY/AGENT INFORMATION:
 NAME: Mirock, S. Leslie
 REGISTRATION NUMBER: 18,872
 REFERENCE/DOCKET NUMBER: 7326-027
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 212-790-9090
 TELEFAX: 212-869-8864
 TELEX: 66141 PENNIE
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1068 amino acids
 TYPE: amino acid
 STRANDEDNESS:
 TOPOLOGY: unknown
 MOLECULE TYPE: protein
 FEATURE:
 NAME/KEY: Human NI (TAN-1)
 LOCATION: 1152...2219
 OTHER INFORMATION: Highly conserved ankyrin repeat
 OTHER INFORMATION: region of No. 6149902ch
 US-09-113-825-2

```

Query Match      5.8%; Score 85; DB 4; Length 1068;
Best Local Similarity 21.6%; Pred. No. 1.5;
Matches 70; Conservative 44; Mismatches 106; Indels 104; Gaps 16;

Qy 11 SIIDEGSFEGLASLALSTSPSAVSORVKALEHHVGRLVSRTPQAKATAGEVLVOAARKM 70
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 758 AVISDFIYQGASL-----HNQTRDTGETALHLAA--YRSRDAAR-----796

Qy 71 VLLQAEYKQAQLSGRLAEPLTIAINADSLSTWFFPVFNVASWGAT-LTI RLDEAHTL 129
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 797 -LLEASADANTODNNGRTPLHAASDAQGVF-----QILIMNRATOLDARMHDGTTPL 849

Qy 130 SLLRRGDVLGAVTREANPVGCEVV-ELG-TMRHLATATPSLRDAYMV-----175
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 850 ILAARLAVEGMLEDLIINSHADVNVDLGKSALHWAANNVN-DAAVVLKNGANKMQN 908

Qy 176 -----DGKLWMAAPVLRFG-----PKDVLQDR-----DLDDGRVD----205
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 909 NREETPLFLAWEGSYETAKVLLDHOFANWDITDHMDRLPRDIAOERMHHDIVRLLDENL 968

Qy 206 -----GPVRRRVSTVP--SAEGF-----GEAIRGLGWGLLPETAQAPMLKA 246
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 969 VRSQLHGAPLGGTPTLSPPCLSPNGYLGSUKPGVQCKVKRPSSKGLACSKRAKD LKA 1028

Qy 247 -----GEVILLDE---IPIDT 259
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

[illegible]

QY 247 -----GEVILLDE---IPIDT 259
|: || |:
Db 2180 WRKKSQDGKGLLDSSCMLSPVDS 2203

Search completed: September 13, 2001, 17:52:12
Job time: 1118 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 13, 2001, 17:52:41 ; Search time 25.14 Seconds
(without alignments)
878.704 Million cell updates/sec

Title: US-09-105-117i-3
Perfect score: 1460
Sequence: 1 MNPQLDITLLSIDGSEFG.....RSLARLTDVDDAAIEGLRP 290
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues
Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_68:.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	538.5	36.9	303	2	G70756
2	410	28.1	300	2	B83100
3	376.5	25.8	297	2	S22098
4	376.5	25.8	297	2	F85947
5	361.5	24.8	298	2	C82318
6	177	12.1	297	2	G83145
7	175	12.0	300	2	D83502
8	174	11.9	295	2	E83302
9	170	11.6	310	2	E83381
10	163.5	11.2	286	2	A83482
11	156	10.7	304	2	E83495
12	152.5	10.4	292	2	G83332
13	152.5	10.4	295	2	D82443
14	151	10.3	302	2	B82036
15	149.5	10.2	339	2	C83290
16	149	10.2	303	2	G83354
17	148	10.1	309	2	S70535
18	143.5	9.8	297	2	D86063
19	143.5	9.8	305	2	F85823
20	142.5	9.8	305	2	D64963
21	141.5	9.7	311	2	H83174
22	140.5	9.6	305	2	B47059
23	140	9.6	312	2	E83624
24	139.5	9.6	297	1	RGEQIY
25	139	9.5	309	2	D64139
26	138.5	9.5	305	2	G82167
27	138.5	9.5	306	2	T35595
28	138.5	9.5	308	2	H65014
29	138	9.5	293	2	H70067

30 136 9.3 294 2 F83023 probable transcript
31 136 9.3 298 2 C83536 probable transcript
32 135 9.2 309 2 C83310 probable transcript
33 133 9.1 309 2 C81954 probable LysR-like
34 131.5 9.0 299 2 E85517 transcription regu
35 131 9.0 309 2 H81011 probable transcript
36 131 9.0 314 2 B83536 probable transcript
37 130 8.9 296 2 E83498 probable transcript
38 130 8.9 317 2 B83158 probable transcript
39 129.5 8.9 298 2 D83319 trans-activator of
40 129.5 8.9 317 2 A36056 regulator for metE
41 129.5 8.9 317 2 D86070 LysR-type transcri
42 129 8.8 292 2 T34997 hypothetical prote
43 128.5 8.8 312 2 F85869 transcription regu
44 128.5 8.8 313 2 A82436 probable transcript
45 128 8.8 302 2 D83630

ALIGNMENTS

RESULT 1

G70756
hypothetical protein Rv1985c - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C:Accession: G70756
R:Colle, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd,
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Sqaes, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno
A:Reference number: A70500; MUID:98295987
A:Accession: G70756
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-303 <COL>
A:Cross-references: GB:Z74025; GB:AL123456; NID:g3261586; PIDN:CAA98410.1; PID:g32615
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: Rv1985c
C:Superfamily: conserved hypothetical protein H11364

Query Match 36.9%; Score 538.5; DB 2; Length 303;
Best Local Similarity 43.1%; Pred. No. 7.8e-36;
Matches 124; Conservative 49; Mismatches 106; Indels 9; Gaps 5;

Qy 5 QLDITLLSIDGSEFGASLALISPSAVSQRVKALEHHVGRVLSVRTOPAKATEGEVLV 64
Db 10 QLAALAAVVELGSEFDAAERLHVTPSAVSQRIKSLQEQGVQVVLVVRKPCRTATTAGIPLL 69
Qy 65 QAARKMVLQAEKTLQSLGR--LAEIPLTIAINADSLSTWFPFVNEVASNGGATLRL 122
Db 70 RLAAQTLAESEALAEAMGNSLKRTRITIAVNADSMATWFSAPVDGL---GDVLDDVRI 126
Qy 123 EDEAHTLLRRGDVLGAVTREAENPACGEVVELGTMTLRLAATPSLRDAYMVDGKLDWA 182
Db 127 EDQHSARLLREGVAMGAVTTERNPVPCRVHPLGEMRYLPVASRPFVQRHLSGFTAA 186
Qy 183 A--MPVRFPGKDVLDQDRDGPVGRRRVSVPSAEGGEGEIRGLGWLGLLPETQA 240
Db 187 AAKAPSLAWNDDGLQDM--LVKKAFFRAITPTTFVPTTEGTTAAARAGLWGMPFEXLA 245
Qy 241 APLMKAGEVILLDEIPDTPTMWRWRLESRLARLTDVAVDAIEGL 288
Db 246 ASPLADGSFVRVCDIHLDPVLMQCKLDSPIIARITD--TVRAASGL 292

RESULT 2

B83100
inhibitor of chromosome initiation Icia PA4363 [imported] - Pseudomonas aeruginosa (S

A;Accession: JN0079
A;Molecule type: DNA
A;Residues: 1-297 <THO>
A;Cross-references: GB:M62865; NID:g146435; PIDN:AAA62780.1; PID:g146436
A;Accession: PS0203
A;Molecule type: protein
A;Residues: 1-28, 'X', 30-35, 'X', 37-38, 'XX', 41 <TH2>
R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;
A;Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A;Title: The complete genome sequence of *Escherichia coli* K-12.
A;Reference number: A64720; MUID:97426617
A;Accession: C65076
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-297 <BIAT>
A;Cross-references: GB:AE000375; GB:U00096; NID:g1789282; PIDN:AACT5953.1; PID:g17892
A;Experimental source: strain K-12, substrain MG1655

```

C:Genetics:
A:Gene: ic1a
A:Map position: 62.8 min
C:Function:
A:Description: binds to the origin (oriC) of Escherichia coli DNA to block initiation
C:Superfamily: regulatory protein ampR
C:Keywords: DNA binding; DNA replication inhibitor
F:21-40/Region: helix-turn-helix motif

Query Match          25.8%; Score 376.5; DB 2; Length 297;
Best Local Similarity 34.1%; Pred. No. 7.3e-23;
Matches 100; Conservative 53; Mismatches 125; Indels 15; Gaps 3;

Qy      6 LDYLLSIIDGSGFEGASLALISPSAVSQRVKALEHHVGRVLVSRTOPAKATAGEVLVQ 65
      | | : : | | : : | : | : | : | : | : | : | : | : | : | : | :
Db      9 LQALDAVIRGGERFAAQKLCITQSAVSQRIKLENNFGQPLLVRTVYPPRTEQGGKLLA 68
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Qy      66 AARKMVLQETRAQLSGRLAETPLTIAINADSLSTWFFPVFNEVASWGGAATLTLRLDE 125
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db      69 LLRQVELLEEELWGDEQGTSTPLLLSLAVNADSLATWLLPALAPVLADSPIRNLQVEDE 128
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Qy      126 AHTLSLLRGDVLGAVTREANPVANGCEVVELGTMRHLAIATPSLRDAYWYDGKLDWRAM- 184
      | | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db      129 TRTQELRRRGVEGWGASIQHOALPSCVLVDKLGALDYLFVSSKPFPAEKYFPNGVTRSALL 188
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :

```

QY 185 - PVLRFGPKDVLQDRDLGRVDGPVGRRRYSIVPSAEGFGEAIRRGLGWGLLPETQAAPM 243
 I : I : I : I : I : I : I : I : I : I : I :
Db 189 APVAFDHLDMHQAIFLQQNFEDLPPGSVPCHVNSSEAFVQLARQGTTCCMPHLQIEKE 248
 I : I : I : I : I : I : I : I : I : I : I :
QY 244 LKAGEVILLDEIPIDTP-----MTWQRWRLESRLARTDAVVDAATGLR 289
 I : ||| : I : I : I : I : I : I : I : I : I :
Db 249 LASGELIDL-----TPGLFRMLYWHRPAPERMMRKVTDALLDYGHKVLR 295
 I : ||| : I : I : I : I : I : I : I : I : I :

RESULT 4
F85947
hypothetical protein icia [imported] - Escherichia coli (strain O157:H7)
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 31-Mar-2001
C:Accession: F85947
R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May
iller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apoda
Nature 409, 529-533, 2001
A:title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: F85947
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-297 <STO>
A:Cross-references: GB:AE005174; NID:g12517445; PIDN:AAG58042.1; GSPDB:GN00145; UWPG:
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: icia

RESULT 6

G83145
probable transcription regulator PA3995 [imported] - Pseudomonas aeruginosa (strain PA01)
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: G83145
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; Lory, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lory, S.; Olson, M.V.
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathogen
A:Reference number: A82950; MUID:20437337
A:Accession: G83145
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-297 <STO>
A:CROSS-references: GB:AE004091; GB:AE004091; PIDN:g9950181; PIDN:AAG07382.1; GSPDB:GN011411
A:Experimental source: strain PA01
C:Genetics:
A:Gene: PA3995
C:Superfamily: hypothetical protein bl875

Query Match 12.1% Score 177; DB 2; Length 297;
Best Local Similarity 28.0%; Pred. No. 7.1e-07;
Matches 78; Conservative 38; Mismatches 119; Indels 44; Gaps 10;

Qy 4 IQDITLSIIDEGSFEGLASLISPSAVSORVKALEHHVGRVLVSR-TOPAKATEAGEV 62
:
Db 14 VLLRTFLEVDSGGFAALADNALTPSAVSGHKIRLEQTACVSLSTRTRRLTQAGET 73

Qy 53 LVQARKNVILQAETHKAQLSRGLAIEPLTINADSLSWTFPPVFNEVASW----- 113

Db 74 LYAYARNIVELERSARALKHGTPIIDRLRIIGASEDFASANLPRLVQRFRKHPEASTLK 133

Qy 114 -GGATLTLRLEDEAHT-----LSLLRRGDVLGAVTREANPVAGCEVVELGTMRLIA-- 165

Db 134 VGITDLLRQAOGRDTDVFGKCRVRVGDDGELLWEERPLYWAAATAIELPAGEPLPALF 193

Qy 166 -TPSL-RDAYMWDGKLDAAMPVLRFGPKVQLQDRDLGDRVDPGVGRRRYSIVPSAGFG 223
:
Db 194 PEPCVYREAAIT--ALGAAPRW-----RLVFESAMAGCLSAALAGFAVTVARSQ--- 243

Qy 224 EAIRRL-----GWGLLPETQ-----AAPMLKAGEVI 250
:
Db 244 --MRGLRELPGEGFTPLPEARFYAFSRQPSLAADALI 280

RESULT 7

D83502
probable transcription regulator PAL141 [imported] - Pseudomonas aeruginosa (strain PA01)
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: D83502
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; Lory, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lory, S.; Olson, M.V.
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathogen
A:Reference number: A82950; MUID:20437337
A:Accession: D83502
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-300 <STO>
A:CROSS-references: GB:AE004544; GB:AE004091; PIDN:g9947060; PIDN:AAG04530.1; GSPDB:GN011411
A:Experimental source: strain PA01
C:Genetics:
A:Gene: PAL141
C:Superfamily: hypothetical protein bl875

Query Match 12.0%; Score 175; DB 2; Length 300;

Best Local Similarity 27.2%; Pred. No. 1e-06;
Matches 82; Conservative 44; Mismatches 121; Indels 54; Gaps 12;

QY 6 LDTLLSIIDEGSFEGLASLSTSPSAVSORVKALEHHVGRVLVSR-TQPAKATEAGEVLV 64
Db 17 LRAIVTVADCGSFTTAAARLHSTQSVSKVRLEEMAGHRLERNDVLTPTDAGDTLL 76

QY 65 QAARKVLLQAEETKAOLSLRLAEIPLTIATINAD-----SLSTWFPVPVNEVASW 113
Db 77 GYARRLLALNDELAEALSGATVLTVRIGVDPDDFAAGRTTERLAAPNRYPQVKLEVTSG 136

QY 114 GGATLRLLEDEAHTLSLRGGDVLGAVTREAN---PVAGCEVEVLGTRMRHLAIAATPSLR 170
Db 137 MSRDLSASVDRGELDVLLKQ-----RRASREALACWPKTCWV-----DSAR 179

QY 171 DAYMVDGKLDWAAMPVLRGPKDVLQDRDLGVRDGPVGR-RVSI VPSA-BGFGEAIRR 228
Db 180 NPC-----IDLPLPIVTPFRGVYRD-EMIAALEA-VGRRWHSITSSLSGLQSLAIAD 232

QY 229 GLWGGL-----LPETQAAPMLKAGEVILLDEIPIDTPMYWQWRLESRLARL 276
Db 233 GMGIGLLPLRAVSAGHAVLPKNVGLPAVDVFEVALLHR-PAADPMVKALARVLSRLAED 291

QY 277 T 277
Db 292 T 292

RESULT 8
E83302
probable transcription regulator PA2758 [imported] - Pseudomonas aeruginosa (strain PA01)
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: E83302
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; Bro-
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho-
gen
A:Reference number: AB2950; MUID:20437337
A:Accession: E83302
A:Molecule type: DNA
A:Residues: 1-295 <STO>
A:Cross-References: GB:AE004703; GB:AE004091; NID:g9948825; PIDN:AAG06146.1; GSPDB:GN001
A:Experimental source: strain PA01
C:Genetics:
A:Gene: PA2758

Query Match 11.9%; Score 174; DB 2; Length 295;
Best Local Similarity 25.3%; Pred. No. 1.2e-06;
Matches 80; Conservative 48; Mismatches 112; Indels 76; Gaps 15;

QY 6 LDTLLSIIDEGSFEGLASLSTSPSAVSORVKALEHHVGRVLVSR-TQPAKATEAGEVLV 64
Db 8 IELFVAVLDGRGSGAARALGRVPSAVSMGIANLEALGFALFERTHREAPRTALARALA 67

QY 65 QAARKVLLQAEETKAOLSLRLAEIPLTIATINAD-----SLSTWFPVPV 107
Db 68 POARMI-----AQQLAQVHALELSQGLSRLSGVADIDSRLFGALRTLSERYPLLD 123

QY 108 NEVASMGATLRLLEDEAHTLSLRGGDV---LGAVTREANP-----VAGCEVEVLGT 158
Db 124 VEVLS-----AAQDDA--LALLHGGRISSLCFGAGISVNPQERFQVGAESLVATIS 173

QY 159 MRHLAIAATPSLRDAYMVDGKLDWAAMPVLRGPKDVLQDRD-LDGRVDPVGRRRVSI 217
Db 174 PRHPALQKPG-QAQLYL-----BELVNVNRQTLVASCDPLADTRPLIAG 215

QY 218 S---AEGFG---EATRRGLGWLGPETQAAPMLKAGEVILLD-----EIPIDTPMYQOR 265
Db 216 ACWRDTSLSGTALEMVEAGIWMGNFPLSRVAPLLATGLRLVRLDFRNTKNELKLPVHAIW-- 273

QY 266 WRLESRLARLTDVAV 281
Db 274 --LKNQPLRAAQELV 287

RESULT 9
E83381
probable transcription regulator PA2123 [imported] - Pseudomonas aeruginosa (strain P
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: E83381
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; Bro-
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa
thogen
A:Reference number: AB2950; MUID:20437337
A:Accession: E83381
A:Molecule type: DNA
A:Residues: 1-310 <STO>
A:Cross-References: GB:AE004639; GB:AE004091; NID:g9948129; PIDN:AAG05511.1; GSPDB:GN
A:Experimental source: strain PA01
C:Genetics:
A:Gene: PA2123
C:Superfamily: conserved hypothetical protein H11364

Query Match 11.6%; Score 170; DB 2; Length 310;
Best Local Similarity 25.7%; Pred. No. 2.7e-06;
Matches 77; Conservative 37; Mismatches 98; Indels 88; Gaps 12;

QY 5 OLDTLISIIDEGSFEGLASLSTSPSAVSORVKALEHHVGRVLVSR-TQPAKATEAGEVLV 64
Db 6 EMTFVSVVDHAFSEARLGTTSQVSRIOQLERRLGRVLLNRLSLDPTORTY 65

QY 65 QAARKV--LLOAET-----KAOLSGRLAEP-----LTIATN 95
Db 66 EHACRLLEQEAASVLDADDLRLGQLRSAPLAFTPRYLAPLLARFAERHPQLRVDVQ 125

QY 96 ADS-LSTWFPVPVNEVASMGATLRLLEDEAHTLSLRGGDVLGAVTREANPVACEV 154
Db 126 ADDRFLNQLQEPDMALRNG-----RLDDSLV-----ARPIT----- 158

QY 155 ELGTRHLAIAATPSLRDAYMVDGKLDWAAMPVLRGPKD-----VLQDRDLGDR--- 203
Db 159 ---ANRHLICASP-----DYLARHGVQPQPEELQAHGDLVYNNRPPGMLCLP 203

QY 204 VDPGVPVRRVSI---VPSAEGFEAIRRGLGWGLLPETQAAPMLKAGEVILLDEIPIDTP 260
Db 204 VDGEPAFRIRVMRMTDSGHQLLEGARAGLGAILLPSFLAADALLAGE--LLPVLPGYSP 261

RESULT 10
A83482
probable transcription regulator PA1309 [imported] - Pseudomonas aeruginosa (strain P
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: A83482
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; Bro-
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa
thogen
A:Reference number: AB2950; MUID:20437337
A:Accession: A83482
A:Molecule type: DNA
A:Residues: 1-286 <STO>
A:Cross-References: GB:AE004560; GB:AE004091; NID:g9947240; PIDN:AAG04698.1; GSPDB:GN
A:Experimental source: strain PA01
C:Genetics:

C:Genetics:

A:Gene: VCA0575

A:Map position: 2

C:Superfamily: probable transcription regulator ybbs

Query Match 10.4%; Score 152.5; DB 2; Length 295;
Best Local Similarity 25.5%; Pred. No. 6.4e-05;
Matches 79; Conservative 62; Mismatches 102; Indels 67; Gaps 17;

QY 1 MNPQLDTLLSIID----EGSFGASIALSISPSAVSORVKALEHHVGRVLSRT--OPAK 55
DB 2 LSPITILEA-LHLIDATERGSGFAAANELNRPSSLSYQIQLEQDLIMIFDRSHRAN 60
QY 56 AYEAGEVLVQAARKMVLQAEYK---AOLSGRLAIPITIAINADSLSTWFFPP---VEN 108
DB 61 FEAGKLIILGR--AILATEKLVNDATLLANGWELDTITIALDG-----IVPAANLFP 112
QY 109 EVASWGATLT-LRLEDE--AHTLSLLR--RGDVLGAVTREPANPVAGCEVVELGTMRLHA 163
DB 113 MVEALGNISKTRVRIQDEILAGWEALATGRADLLICPRIEALP-QDVKAETIGTKMIW 171
QY 164 IATPS-----LRDAY----WYDGKLDWAAMPVLRFGPKDVLQDRDLGRVDG 206
DB 172 VAAPTHYVHRSGEFNEAREKRYATAIADTAREQPMASV-----NILQ----- 215
QY 207 PVGRRRSVSPSAEGFGEAIRRGLGMLLPETQAAAPMLKAGEVILL---DEIPIDTPMYW 263
DB 216 ---RQPLTVSNLDAKCKALVAGLGIGTLPLOVAQPYIDKGLKAIHGSDELMDIVLAN 272
QY 264 QRWRL-ESRS 272
DB 273 RRNQMGAKS 282

RESULT 14

B82036

transcription regulator lysr family VC2760 [imported] - Vibrio cholerae (strain N16961)

C:Species: Vibrio cholerae

C>Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001

C:Accession: B82036

R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.; Chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragol, I.; Sellers, B. I., R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.

Nature 406, 477-483, 2000

A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.

A:Reference number: A82035; MUID:20406833

A:Accession: B82036

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-302 <HEI>

A:Cross-references: GB:AE004342; GB:AE003852; NID:g957358; PIDN:AAF95899.1; GSPDB:GN001

A:Experimental source: serogroup O1; strain N16961; biotype El Tor

C:Genetics:

A:Gene: VC2760

A:Map position: 1

C:Superfamily: probable transcription regulator ybbs

Query Match 10.3%; Score 151; DB 2; Length 302;
Best Local Similarity 26.1%; Pred. No. 8.8e-05;
Matches 69; Conservative 53; Mismatches 118; Indels 24; Gaps 10;

QY 6 LDTLLSIIDEGSFGASIALSISPSAVSORVKALEHHVGRVLSR--TOPAKATEAGEVLV 64
DB 8 LEMLDVARLSGFTAAAEVLHVPAISVSRQISELGVLFRLPRKVELTPAGELFM 67
QY 65 ----QAARKMVLQAEYKQISGRLAELPTIAINADSLSTWFFPPVFNVEASWGATLTL 120
DB 68 LEAROLLQMEIRIAQTKRAAGWKRKLKVTLD--NVVKLDKMPWEAFYQTFDFAEIQI 126
QY 121 RLEDEAHTLSLLRGD---VLGAVTREPANPVAG-CEVVELGTMRLHAITATPS---LRDAY 173

DB 127 NMEVFNWSWEAIAQGRADIVIGATA--AVPVGSGDFEVRDMGILDWAFVMSPNHPCVREQN 184
QY 174 MYDGKLDWAAMPVLRFGPKDVLQDRDLGRVDGPGVRRRVSVIPSAEGFGEAIRRGLGWS 233
DB 185-LSEGFTS-QFLAICLDTSVLPKRHTE-----HYPKQRLLLPNWSAIESLSKSLGIVG 238
QY 234 LLPETOAAAPMLKAG---EVILLDE 254
DB 239 YMPRHMAQPLLASGQLVEXVLPDE 262

RESULT 15

C83290

probable transcription regulator PA2848 [imported] - Pseudomonas aeruginosa (strain P)

C:Species: Pseudomonas aeruginosa

C>Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000

C:Accession: C83290

R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L

; Lory, S.; Olson, M.V.

Nature 406, 959-964, 2000

A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa

A:Reference number: A82950; MUID:20437337

A:Accession: C83290

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-339 <STO>

A:Cross-references: GB:AE004711; GB:AE004091; NID:g9948927; PIDN:AAG06236.1; GSPDB:GN

A:Experimental source: strain PA01

C:Genetics:

A:Gene: PA2848

C:Superfamily: conserved hypothetical protein H11364

Query Match 10.2%; Score 149.5; DB 2; Length 339;
Best Local Similarity 27.2%; Pred. No. 0.00013;
Matches 83; Conservative 40; Mismatches 131; Indels 51; Gaps 13;

QY 7 DPL-----LSIIDEGSFGASIALSISPSAVSORVKALEHHVGRVLSRTQPA-KATE 58
DB 25 DTLSEQFVLYLDVLDGGSFSAARKHPLTPSAVARRMDALERAVGSTLLVTRTHAVRATP 84
QY 59 ACEVLVQAARKW-----LLOATKALQSG--RLAEIPLTIANADSLSTWFFPPVFNVEAS 112
DB 85 AGLAFADRARRIVTELRLARAEAVSLSTAPOGLIRIDAPVPGRRHLA---PAVADFLKA 141
QY 113 WGGATLTLRLEDEAHTLSLLRRGDVLGAV--TREANPVAGCEVW--ELGTMRLHAITATPS 168
DB 142 NPGLDVQLRLIGSFIDL-----QGEHLGEVDLVLRAGPULPDSRLVATSLAPMVVVVCASPE 197
QY 169 LRDAYMVDGK-----LDWAAMP---VLRFGPKDVLQDRDLGRVDGPGVRRRVSV 216
DB 198 YLREHGVSPASPGELPEHAGIDHNDLSPYAWRF-----QHDGKLOHLRPRKARLAT 248
QY 217 PSAEGFGEAIRRGLGMLLPETQAAAPMLKAGEVILL---DEIPIDTP--MYQWRWLESR 271
DB 249 NNAEAMWDAALAGLGLIAHLPTWLCSEYLLRGELQALFCDGDLPAAEPTCIYALRLEREAS 308
QY 272 SLARL 276
DB 309 SRTLL 313

Search completed: September 13, 2001, 17:52:42

Job time: 1067 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.
OM protein - protein search, using sw model
Run on: September 13, 2001, 17:57:09 ; Search time 15.26 Seconds
(without alignments)
650.989 Million cell updates/sec
Title: US-09-105-117i-3
Perfect score: 1460
Sequence: 1 MNPQLDITLLSIIDEGSFEG.....RSLARLTDVVDAIEGLRP 290
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5
Searched: 93435 seqs, 34255486 residues
Total number of hits satisfying chosen parameters: 93435
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries
Database : SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1460	100.0	290	1	LYSG_CORGL
2	538.5	36.9	303	1	YJ85_MYCTU
3	376.5	25.8	297	1	ICIA_ECOLI
4	365.5	25.0	297	1	ICIA_EDWIC
5	355.5	24.3	299	1	ICIA_AERSA
6	148	10.1	309	1	LTRA_KLEPN
7	142.5	9.8	305	1	NAC_ECOLI
8	140.5	9.6	305	1	NAC_KLEAE
9	139.5	9.6	297	1	ILYV_ECOLI
10	139	9.5	309	1	MYTR_HAEIN
11	138.5	9.5	308	1	YFER_ECOLI
12	136.5	9.3	317	1	METR_SALTY
13	136	9.3	294	1	LRRA_SYNP7
14	129.5	8.9	317	1	MYTR_ECOLI
15	128	8.8	295	1	TEPR_ALCEU
16	127.5	8.7	301	1	ESTR_ACTICA
17	127.5	8.7	311	1	OXRY_MYCLE
18	127.5	8.7	312	1	LRHA_ECOLI
19	127.5	8.7	326	1	BLAA_STRCI
20	126.5	8.7	309	1	NTCB_SYNP7
21	126.5	8.7	317	1	CFXR_ALCEU
22	126	8.6	316	1	PECT_ERWCH
23	125.5	8.6	317	1	YC30_PORPU
24	121	8.3	310	1	YDHB_ECOLI
25	117.5	8.0	298	1	TRPI_PSESY
26	116.5	8.0	306	1	OPRR_PSEAE
27	115.5	7.9	298	1	CBRR_RHURU
28	115.5	7.9	311	1	OXTR_MYCAV
29	111.5	7.6	307	1	YEAT_ECOLI
30	111	7.6	292	1	HVRB_RHOCA
31	111	7.6	299	1	YCJZ_ECOLI
32	108.5	7.4	308	1	YBSB_ECOLI
33	108	7.4	312	1	YMR2_MYCTU

34	107.5	7.4	298	1	YHAI_ECOLI	P42623	escherichia
35	107.5	7.4	309	1	NTCB_SYNP3	P74422	synecocyst
36	107	7.3	292	1	ILYV_HAEIN	P44821	haemophilus
37	107	7.3	298	1	Y4WQ_RHISN	P55576	rhizobium s
38	107	7.3	304	1	TU44_AGRVI	P52669	agrobacteri
39	106.5	7.3	293	1	YKUM_BACSU	O34827	bacillus su
40	106.5	7.3	294	1	CLCR_PSEPU	Q05840	pseudomonas
41	106.5	7.3	299	1	YWFK_BACSU	P39847	bacillus su
42	106.5	7.3	308	1	YTLI_BACSU	O35038	bacillus su
43	106.5	7.3	310	1	CBRR_RHOSH	P52690	rhodobacter
44	105.5	7.2	143	1	BUDR_ENTAE	P52665	enterobacte
45	104.5	7.2	299	1	CYNR_ECOLI	P27111	escherichia

ALIGNMENTS

RESULT 1	LYSG_CORGL	STANDARD;	PRT;	290 AA.
AC	P94632;			
DT	15-DEC-1998 (Rel. 37, Created)			
DT	15-DEC-1998 (Rel. 37, Last sequence update)			
DT	15-DEC-1998 (Rel. 37, Last annotation update)			
DE	LYSINE EXPORT TRANSCRIPTIONAL REGULATORY PROTEIN LYSG.			
GN	LYSG.			
OS	Corynebacterium glutamicum (Brevibacterium flavum).			
OC	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;			
OC	Actinomycetales; Corynebacterineae; Corynebacteriaceae;			
OC	Corynebacterium.			
OX	NCBI_TaxID=1718;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=R127;			
RX	MEDLINE=97126810; PubMed=8971704;			
RA	Vrljic M., Sahm H., Eggeling L.;			
RT	"A new type of transporter with a new type of cellular function: L-			
RT	lysine export from Corynebacterium glutamicum."			
RL	Mol. Microbiol. 22:815-826(1996).			
CC	-!- FUNCTION: PROBABLE TRANSCRIPTIONAL REGULATOR OF LYSG.			
CC	-!- SIMILARITY: BELONGS TO THE LYSR FAMILY OF TRANSCRIPTIONAL			
CC	REGULATORS.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement. (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
CC	EMBL; X96471; CAA65323.1; -			
DR	InterPro; IPR000847; -			
DR	Pfam; PF00126; HTH_1; 1			
DR	PROSITE; PS00044; HTH_LYSR_FAMILY; 1.			
KW	DNA-binding; Transcription regulation.			
FT	DNA_BIND 18 37 H-T-H MOTIF (POTENTIAL).			
SQ	SEQUENCE 290 AA; 31388 MW; 9CE0A63F775FCB74 CRC64;			

Query Match 100.0%; Score 1460; DB 1; Length 290;
Best Local Similarity 100.0%; Pred. No. 6.4e-110;
Matches 290; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MNPQLDITLLSIIDEGSFEGASLALSI	SPSAVSQRVKALEHHVGRVLVSRTPAKATEAG	60
Db	1	MNPQLDITLLSIIDEGSFEGASLALSI	SPSAVSQRVKALEHHVGRVLVSRTPAKATEAG	60
QY	61	EVLVQAARKMVLQAEETKAQLSGRLAE	IPITIAINADSLSTWPPVFNVEASGGATLTL	120
Db	61	EVLVQAARKMVLQAEETKAQLSGRLAE	IPITIAINADSLSTWPPVFNVEASGGATLTL	120
QY	121	RLEDEAHTLSLLRRGDVLGAVTREPAN	PVACEVVEIGTMRLHAIATPSLRDAVWDGKLD	180

```

Db 121 RLEDEAHTLSLLRRGDVLGAVTREPANPVAGCEVVELGTMRLHAIATPSLRDAYVMDGKLD 180
Qy 181 WAAMPVLRFGPKDVLQDRDLGRVDPGVRRRVSVPAEGFGEAIRRGLGWLGPETQA 240
Db 181 WAAMPVLRFGPKDVLQDRDLGRVDPGVRRRVSVPAEGFGEAIRRGLGWLGPETQA 240
Qy 241 APMKAGEVILLDEIPIDTPMYQWRWRLESRLARTDAVVDAAIEGLRP 290
Db 241 APMKAGEVILLDEIPIDTPMYQWRWRLESRLARTDAVVDAAIEGLRP 290.

RESULT 2
ID YJ85_MYCTU STANDARD; PRT; 303 AA.
AC Q10872;
DT 01-OCT-1996 (Rel. 34, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE HYPOTHETICAL TRANSCRIPTIONAL REGULATOR RV1985C.
GN RV1985C OR MTCY39.34.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98259987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeyer K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squires R., Sulston J.E.,
RA Taylor K., Whitehead S., Barrall B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
CC -1- SIMILARITY: BELONGS TO THE LYSR FAMILY OF TRANSCRIPTIONAL
CC REGULATORS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z74025; CAA98410.1;
CC TuberculList; RV1985c;
CC InterPro: IPR000847;
CC Pfam: PF00126; HTH_L1; 1.
CC PRINTS; PR00039; HTHLYSR.
CC PROSITE; PS00044; HTH_LYSR_FAMILY; 1.
CC Hypothetical protein; Transcription regulation; DNA-binding.
FT DNA_BIND 23 42 H-T-H MOTIF (POTENTIAL)
SQ SEQUENCE 303 AA; 32836 MW; 344BDB1056207144 CRC64;

Query Match 36.9%; Score 538.5; DB 1; Length 303;
Best Local Similarity 43.1%; Pred. No. 3.8e-36;
Matches 124; Conservative 49; Mismatches 106; Indels 9; Gaps 5;

Qy 5 QLDTLLSIDGSEFEGASLALISPSAVSQRVKALEHHVGRVLSRTPAKATAGEVLV 64
Db 10 QLAALAAVVLGSEFEGASLALISPSAVSQRVKALEHHVGRVLSRTPAKATAGEVLV 69
Qy 65 QAARKVWLQATTAQLSGR--LAEIPLTIAINADSLSTFPFPVNEVASWGATLTLLRL 122
Db 70 RLAAQTALLESEALAEAGNGLASLAKRTRITIAVNDASMTATWFSVFDGL---GDVLLDVRI 126

```

```

Qy 123 EDEAHTLSLLRRGDVLGAVTREPANPVAGCEVVELGTMRLHAIATPSLRDAYVMDGKLDWA 182
Db 127 EDDHSARLLREGVANGAVTTERNPVPCGRVHPLCEMRYLPVASRPFVQRHLSDFGTAA 186
Qy 183 A--MPVLRFGPKDVLQDRDLGRVDPGVRRRVSVPAEGFGEAIRRGLGWLGPETQA 240
Db 187 AAKAPSLAWNDDGLQDM-LVRKAFRRATITRPTHEVPTTEGFTAAARAGLWGMFPEKLA 245
Qy 241 APMKAGEVILLDEIPIDTPMYQWRWRLESRLARTDAVVDAAIEGL 288
Db 246 ASPLADGSGFVRVCDIHLDPVLYWCQKLDSPILARIITD-TVRAAASGL 292

RESULT 3
ID ICIA_ECOLI STANDARD; PRT; 297 AA.
AC P24194;
DT 01-MAR-1992 (Rel. 21, Created)
DT 01-MAR-1992 (Rel. 21, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE CHROMOSOME INITIATION INHIBITOR (ORIC REPLICATION INHIBITOR).
GN ICIA.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-41.
RC STRAIN=K12 / W3110;
RX MEDLINE=91239508; PubMed=2034653;
RA Theony B., Hwang D.S., Fradkin L., Kornberg A.;
RT "icia" an Escherichia coli gene encoding a specific inhibitor of
RT chromosomal initiation of replication in vitro.;
RL Proc. Natl. Acad. Sci. U.S.A. 88:4066-4070(1991).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RA Roy I., Leadlay P.F.;
RL Submitted (JUN-1992) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [4]
RP SEQUENCE OF 1-176 FROM N.A.
RC STRAIN=K12;
RX MEDLINE=93374859; PubMed=8366047;
RA Hove-Jensen B., Møgelgaard M.;
RT "Escherichia coli rpiA gene encoding ribose phosphate isomerase A.";
RL J. Bacteriol. 175:5628-5635(1993).
CC -1- FUNCTION: SPECIFIC INHIBITOR OF CHROMOSOMAL INITIATION OF
CC REPLICATION IN VITRO. BINDS THE THREE 13-MERS IN THE ORIGIN
CC (ORIC) TO BLOCK INITIATION OF REPLICATION.
CC -1- SUBUNIT: BEHAVES AS HOMODIMER IN SOLUTION.
CC -1- SIMILARITY: BELONGS TO THE LYSR FAMILY OF TRANSCRIPTIONAL
CC REGULATORS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M62865; AAA62780.1;

```

```

DR EMBL; AF037440; AAB92569.1; -.
DR InterPro; IPR000847; -.
DR Pfam; PF00126; HTH_1; 1.
DR PRINTS; PR00039; HTHLYSR.
DR PROSITE; PS00044; HTH_LYSR_FAMILY; 1.
KW DNA-binding; Transcription regulation.
FT DNA_BIND 21 40 H-T MOTIF (BY SIMILARITY).
SQ SEQUENCE 297 AA; 33392 MW; 93BEC456B2419989 CRC64;

Query Match          25.08; Score 365.5; DB:1; Length 297;
Best Local Similarity 33.48; Pred. No. 2.6e-22;
Matches 98; Conservative 51; Mismatches 129; Indels 15; Gaps 3;

Qy 6 LDTLLSIIDEGSFAGSLALSISPSAVSORVKALEHHVGRVLVSRTOPAKATEAGEVLVQ 65
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 9 LQALDAVIRERGFERRAQKLCITOSAVSORIKOLENLFQGPLLVRTVPPTPEQGORLLA 68
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Qy 66 AARKMVLLOAETKAQLSGRLAEIPLTITAINADLSLTWFPFVENVASGGATLTLRLDEE 125
   :: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 69 LLHQVELLEEEELGNDNSDGPPLLSTLAVNADSLATLWPALKPKYLDSPRLNLQVEDE 128
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Qy 126 AHTLSLRRGDVLGAVTREANPVAGCEVVELGTMRHLAIATPSLRDAYMVDS--KLDWAA 183
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 129 SRTQERLARGEVGAVSLQPQLPCLVDRLGALDYLCASFAPFAARYFPNGVTSSLLK 188
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Qy 184 MPVLRFPGKVQLQRDLGRRGVGRRRVSTIVPSAEGFGBAIRGLGWGLLPETQQAAPM 243
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 189 APAVAFDHLDHMQAFLQONFELPGSPVPCHTVNSSEAFVQLAQLGTTCCMLPHLQVERE 248
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Qy 244 LKAGEVILLDEIPIDTP-----MYQWRWLERSLARLTDAVVDAIEGLR 289
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 249 LRKGVLIDL-----TPGMVORMLYHRFPAPESLMRRVRVTDALISHGRDVLR 295
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 5
ICIA_AERSA
ID ICIA_AERSA STANDARD; PRT; 299 AA.
AC P70773;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE CHROMOSOME INITIATION INHIBITOR (ORIC REPLICATION INHIBITOR).
GN ICIA.
OS Aeromonas salmonicida.
OC Bacteria; Proteobacteria; gamma subdivision; Aeromonadaceae;
OC Aeromonas.
OX NCBI_TaxId=645;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCIMB 1102;
RA Swift S., Karylshev A.V., Fish L., Durant E.L., Winson M.K.,
RA Williams P., MacIntyre S., Stewart G.S.A.B.;
RL Submitted (SEP-1996) to the EMBL/GenBank/DDBJ databases.
CC -!- FUNCTION: SPECIFIC INHIBITOR OF CHROMOSOMAL INITIATION OF
CC REPLICATION IN VITRO. BINDS THE THREE 13-MERS IN THE ORIGIN
CC (ORIC) TO BLOCK INITIATION OF REPLICATION (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE LYSR FAMILY OF TRANSCRIPTIONAL
CC REGULATORS.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U65741; AAB70016.1; -.
DR InterPro; IPR000847; -.
DR Pfam; PF00126; HTH_1; 1.
DR PRINTS; PR00039; HTHLYSR.
DR PROSITE; PS00044; HTH_LYSR_FAMILY; 1.
```

KW DNA-binding; Transcription regulation.
 FT DNA_BIND 21 40 H-T-H MOTIF (POTENTIAL).
 SQ SEQUENCE 299 AA; 33410 MW; 5EBEBCA44CC5A68D CRC64;

Query Match 24.3%; Score 355.5; DB 1; Length 299;
 Best Local Similarity 33.6%; Pred. No. 1.7e-21;
 Matches 98; Conservative 49; Mismatches 126; Indels 19; Gaps 6;

Qy 11 SIIDEGSFEGLASLSPSAVSQRYKALEHHVGRVLSRTPQAKATEAGEVLVQAARKM 70
 Db 14 AVMOQFNERAARLHTQSAISRIKLEQOFAEPLLRISQPLQATPLGOKLLAHYQV 73
 Qy 71 VLLQAEKALSGRL-----AEPLTAINADSLSTWFFPVFNEVASWGGATLTRED 124
 Db 74 RQLELE-----LAGEIAPDEPOAPITVSTAVNADSLATWFLPALAPLLEHQHPIELNLLVDD 129
 Qy 125 EAHTLSLLRRGDLGAVTREPANVAGCEVVELGTMRLHAIATPSLRDAYMVDGKLDNAAM 184
 Db 130 ECRITLDVRGQAFAGVSLHGOPLAGCCVDELGEMRYLLTASPAFVARHFPAG-LTPAAL 188
 Qy 185 ---PVLREGPKDVLQDRDLGCRVGVRRRVSIVPSAEGGEAIRRGLGHLGLPETQAA 241
 Db 189 AKTPAFAEDQDRDHVSPMARHFGLEPGCYCHTVRSSEAFVMAEQGLAYCLIPQLQIR 248
 Qy 242 PMLKAGEVILLDEIP---IDTPMYWQWRLESRLARLTDAVDAAIETGLRP 290
 Db 249 QQLAOG---ILLDLSPSHLIEFLYHWRVLERGLHKLQISQRLISEGRALQP 298

RESULT 6
 LTRA_KLEPN
 ID LTRA_KLEPN STANDARD; PRT; 309 AA.
 AC P52689;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE PROBABLE TRANSCRIPTIONAL REGULATOR LTRA.
 GN LTRA.
 OS Klebsiella pneumoniae.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Klebsiella.
 OX NCBI_TaxID=573;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 13882;
 RX MEDLINE=96342382; PubMed=8748036;
 RA Bott M., Meyer M., Dimroth P.;
 RT "Regulation of anaerobic citrate metabolism in Klebsiella pneumoniae";
 RL Mol. Microbiol. 18:533-546(1995).
 CC -1- SIMILARITY: BELONGS TO THE LYSR FAMILY OF TRANSCRIPTIONAL REGULATORS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC DR EMBL; U31464; AAC44735.1; -;
 CC DR InterPro; IPR000847; -;
 CC DR Pfam; PF00126; HTH_1; 1.
 CC DR PROSITE; PS00044; HTH_LYSR_FAMILY; 1.
 KW Transcription regulation; DNA-binding.
 FT DNA_BIND 21 40 H-T-H MOTIF (POTENTIAL).
 SQ SEQUENCE 309 AA; 34357 MW; 5C54D416605C734A CRC64;

Query Match 10.1%; Score 148; DB 1; Length 309;
 Best Local Similarity 26.3%; Pred. No. 7e-05;

Matches 80; Conservative 50; Mismatches 126; Indels 48; Gaps 15;

Qy 6 LDTLLSIIDEGSFEGLASLSPSAVSQRYKALEHHVGRVLSRTPQAKATEAGEVLV 64
 Db 9 LALFQIVDQGSFSAVARQNGITPSAVSRVSRLEREMGCKLLQRTTRKRLSDAGETIY 68
 Qy 65 QAARKMV---LQAEKALSGRLAEIPLTIAI-NADSLSTWFFPVFNEVASWGGATLTLR 121
 Db 69 QHAQOMLEAARQAMDSAGSRQTVAGGKLTLSVPKAVGFRVTHPLMMAFFHRYPOVDVCLR 128
 Qy 122 LEDAHTLSLLRRG-DVLGAVTREPANVAGCEVVELGTMRLHAIATPSLRDAYMVDGKLD 180
 Db 129 LED--RPLDFDDGDLALRITDTPSP--GLHGKPLMPIRHVICAT----EAY-----LQ 175
 Qy 181 WAAMPVLRFGPKDV-----LQDRDLGCRVGVRRRVSIVPSAEGG----- 223
 Db 176 QHGTP---YTPQDLRAHSCISLGETPADARK----FRREGKTTVQTYRYAANHIAVR 228
 Qy 224 -EATRRGLGWLGLPETQAAPMLKAGEVI-LLDE---IPIDTPMYWQWRLESRLARLTLD 278
 Db 229 LDVNRQHLGIGSLPLFTAREALANGDIVQLPEWFEFISYSGDLWLLWAGDKHMPARMR- 287

RESULT 7
 NAC_ECOLI
 ID NAC_ECOLI STANDARD; PRT; 305 AA.
 AC Q47005;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE NITROGEN ASSIMILATION REGULATORY PROTEIN NAC (NITROGEN ASSIMILATION CONTROL PROTEIN).
 GN NAC.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12;
 RX MEDLINE=98155143; PubMed=9495755;
 RA Muse W.B., Bender R.A.;
 RT "The nac (nitrogen assimilation control) gene from Escherichia coli.";
 RL J. Bacteriol. 180:1166-1173(1998).
 CC [2]
 CC SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of Escherichia coli K-12.";
 RL Science 277:1453-1474(1997).
 CC [3]
 CC SEQUENCE FROM N.A.
 RC STRAIN=K12;
 RX MEDLINE=97251358; PubMed=9097040;
 RA Itoh T., Aiba H., Baba T., Fujita K., Hayashi K., Inada T.,
 RA Isono K., Kasai H., Kimura S., Kitakawa M., Kitagawa M.,
 RA Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
 RA Nakade S., Nakamura Y., Nishimoto H., Nishio Y., Oshima T.,
 RA Saito N., Sampei G., Seki Y., Sivasubram S., Tagami H.,
 RA Takeda J., Takemoto K., Wada C., Yamamoto Y., Horiuchi T.;
 RT "A 460-kb DNA sequence of the Escherichia coli K-12 genome
 RL corresponding to the 40.1-50.0 min region on the linkage map.";
 CC DNA Res. 3:379-392(1996).
 CC -1- FUNCTION: TRANSCRIPTIONAL ACTIVATOR FOR THE HUT, PUT AND URE
 CC OPERONS AND REPRESSOR FOR THE GDH AND GLTB OPERONS IN RESPONSE TO

CC NITROGEN LIMITATION. NEGATIVE REGULATOR OF ITS OWN EXPRESSION (BY
 CC SIMILARITY). BELONGS TO THE LYSR FAMILY OF TRANSCRIPTIONAL
 CC REGULATORS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; U56736; AAD14869.1; -;
 CC EMBL; A6002290; AAC75050.1; -;
 CC EMBL; D90837; BAA15806.1; -;
 CC EcoGene; EGI4265; nac.
 CC InterPro; IPR000847; -;
 CC Pfam; PF00126; HTH_1; 1.
 CC PRINTS; PR00039; HTHLYSR.
 CC PROSITE; PS00044; HTH_LYSR_FAMILY; 1.
 CC Transcription regulation; DNA-binding; Activator; Repressor;
 KW Nitrate assimilation.
 FT DNA_BIND 18 37 H-T-H MOTIF (POTENTIAL).
 SQ SEQUENCE 305 AA; 32835 MW; B0B5DC408C6C6821 CRC64;

Query Match 9.8%; Score 142.5; DB 1; Length 305;
 Best Local Similarity 25.3%; Pred. No. 0.00019;
 Matches 75; Conservative 45; Mismatches 78; Indels 99; Gaps 14;
 QY 1 MNPQIDLLSIIDSEFGASLALSISPSAVSQRVKALEHHVGRVLSRT-OPAKATEA 59
 DB 1 MNPFRLLKYFKYKIDIGSLTQAQAEVLHIAQPALSQQVATLEGELNQQLLIRTKRGVTPDA 60
 QY 60 GEVLVQAARKWVLQAEKRAQ-----LSGRLA-----EPLTIAINADS 98
 DB 61 GKLYTHAR-AIRQCE-QQLAVHNVGQALSGQVSIIGFAPGTAASITMPLLOAVRAE- 117
 QY 99 LSTWFFPPVFEVNASWGATLTLRLEDEAHTLSLLRRGDVLGAVTREANPVAG----- 150
 DB 118 ----PFEIYIHLHNSGAVLNKELIN--HOLDM-----AVIYHSPVAGVSSQALLK 163
 QY 151 -----C--EVEVLGTMRLHATATSLRDYAVWDGKLDWAAMPVLRFGPKDVLQDR 198
 DB 164 EDLFLVGTQDCPGQSDVNAIAQMNLFPS-----DYSAIRL----- 200
 QY 199 DLDGRYDGPVGRBRVSISAEFGFE-----AIRCLGWGLLPETQAAPMLKA 246
 DB 201 ----RVDEAFSLRL----TAKVIGEIESITLTAIAASGMGAVLPESAARSLCGA 249

RESULT 8
 ID NAC_KLEAE STANDARD; PRT; 305 AA.
 AC Q08597;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE NITROGEN ASSIMILATION REGULATORY PROTEIN NAC (NITROGEN ASSIMILATION
 DE CONTROL PROTEIN).
 GN NAC.
 OS Klebsiella aerogenes.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Klebsiella.
 OX NCBI_TaxID=28451;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=W70 / KC1043;
 RX MEDLINE=93209957; PubMed=8458853;
 RA Schwacha A., Bender R.A.;
 RT "The nac (nitrogen assimilation control) gene from Klebsiella
 RT aerogenes.";

RL J. Bacteriol. 175:2107-2115(1993).
 CC -!- FUNCTION: TRANSCRIPTIONAL ACTIVATOR FOR THE HUT, PUT AND URE
 CC OPERONS AND REPRESSOR FOR THE GDH AND GLUT OPERONS IN RESPONSE TO
 CC NITROGEN LIMITATION. NEGATIVE REGULATOR OF ITS OWN EXPRESSION.
 CC -!- SIMILARITY: BELONGS TO THE LYSR FAMILY OF TRANSCRIPTIONAL
 CC REGULATORS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; L01114; AAA18173.1; -;
 CC PIR; B47099; B47099.
 CC InterPro; IPR000847; -;
 CC Pfam; PF00126; HTH_1; 1.
 CC PRINTS; PR00039; HTHLYSR.
 CC PROSITE; PS00044; HTH_LYSR_FAMILY; 1.
 CC Transcription regulation; DNA-binding; Activator; Repressor;
 KW Nitrate assimilation.
 FT DNA_BIND 18 37 H-T-H MOTIF (POTENTIAL).
 SQ SEQUENCE 305 AA; 32754 MW; 775F219AF32AFADC CRC64;

Query Match 9.6%; Score 140.5; DB 1; Length 305;
 Best Local Similarity 28.1%; Pred. No. 0.00028;
 Matches 77; Conservative 39; Mismatches 105; Indels 53; Gaps 12;
 QY 1 MNPQIDLLSIIDSEFGASLALSISPSAVSQRVKALEHHVGRVLSRT-OPAKATEA 59
 DB 1 MNPFRLLKYFKYKIDIGSLTQAQAEVLHIAQPALSQQVATLEGELNQQLLIRTKRGVTPTEA 60
 QY 60 GEVLVQAARKWVLQAEKRAQ-----TKAOLSGRLAEIPLTIAINADSLSTWPPVF 107
 DB 61 GKLYTHAR-TILRQCEQAQLAVNVGQTLRGQVSIIGLPGTAASAITWPLLOT-----VR 115
 QY 108 NEVASWGGATLTLRLEDEAHTL--SLLRGDVLGAVTREANPVAGVEVLGTMRLHIA 165
 DB 116 NELPE-----VMVYLQESSCTALNDKLLAGQLDMAVLYERSPVAG-----IVS 158
 QY 166 TPSLR-DAYMVDGK-----LDWAAMPVLR-FGPKDVLQDRDLDGRYDGPVGRBRVS 215
 DB 159 QPLLKEDLYLVGTRDCPGQSDVLTAVAEMNLFPLRDYSAVR---ARVTEAFTLRRLSAKI 215
 QY 216 ---VPSAEFGFAIRRCGLGWGLLPETQAAPMLKA 246
 DB 216 IGEIESITLTAIAASGMGATVLPESAARSLCGA 249

RESULT 9
 ID ILVY_ECOLI STANDARD; PRT; 297 AA.
 AC P05827;
 DT 01-NOV-1988 (Rel. 09, Created)
 DT 01-NOV-1988 (Rel. 09, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE TRANSCRIPTIONAL ACTIVATOR PROTEIN ILVY.
 GN ILVY.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12;
 RX MEDLINE=86111952; PubMed=3003115;
 RA Wek R.C., Hatfield G.W.;
 RT "Nucleotide sequence and in vivo expression of the ilvY and ilvC
 RT genes in Escherichia coli K12. Transcription from divergent
 RT overlapping promoters.";


```

QY 164 IATPSLRDAYMVGKLD--WAAMPVLRGPKDVLDQRDLGRV---DGPVGRRRVSIVPS 218
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 179 --SLAEKDWEAEFDIDETWYTPV---PDDML---DLWRKVLKSKGINTPRTTTELT 229
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 219 AEGFGEARRGLGWLGPETQAAPMLKAGEVI 250
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 230 AMIQVASRRGI--ATIPYWAALPYLEKGYV 259
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 11
YF_ER_ECOLI
ID YF_ER_ECOLI STANDARD; PRT; 308 AA.
AC P77500;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE HYPOTHETICAL TRANSCRIPTIONAL REGULATOR IN XAPA-LIG INTERGENIC REGION.
GN YF_ER.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-K12;
RX MEDLINE=97349980; PubMed=9205837;
RA Yamamoto Y., Alba H., Baba T., Hayashi K., Inada T., Isono K.,
RA Itoh T., Kimura S., Kitagawa M., Makino K., Miki T., Mitsuhashi N.,
RA Mizobuchi K., Mori H., Nakade S., Nakamura Y., Nashimoto H.,
RA Oshima T., Oyama S., Saito N., Sampei G., Satoh Y., Sivasubram S.,
RA Tagami H., Takahashi H., Takeda J., Takemoto K., Uehara K., Wada C.,
RA Yamagata S., Horiuchi T.;
RT "Construction of a contiguous 874-kb sequence of the Escherichia coli
RT K-12 genome corresponding to 50.0-68.8 min on the linkage map and
RT analysis of its sequence features."
RL DNA Res. 4:91-113(1997).
CC -1- SIMILARITY: BELONGS TO THE LYSR FAMILY OF TRANSCRIPTIONAL
CC REGULATORS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE000328; BAA75462.1; -
CC EMBL; D90870; BAA16280.1; -
CC EcoGene; EG14159; yfer.
CC InterPro; IPR000847; -
CC Pfam; PF00126; HTH_1; 1.
CC PRINTS; PR00039; HTHLYSR.
CC PROSITE; PS00044; HTH_LYSR_FAMILY; 1.
KW Hypothetical protein; Transcription regulation; DNA-binding.
FT DNA_BIND 20 39 H-T-H MOTIF (POTENTIAL).
SQ SEQUENCE 308 AA; 33903 MW; 23B6A771E4AB587 CRC64;

Query Match 9.5%; Score 138.5; DB 1; Length 308;
Best Local Similarity 25.1%; Pred. No. 0.0004;
Matches 78; Conservative 51; Mismatches 119; Indels 63; Gaps 18;

```

```

QY 5 QLOTLTSLIDEGSEFASLALISPSAVSORVKALEHHVG-RVLVSTQPAKATEACEVL 63
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 7 QLKVFVVAQEKSFGRAGERIGLSQSAVSHSVKLEHNHTGVRLDRTTREVLTDAQ-- 64
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 64 VQAAKKVLLQAKTKA-----OLSGRLAEIPLTIAINADSLSTWPP--VFNEVA 111
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 65 -QLALRLERLDELNLTDRGTGRMGQQLSGKVR-----VAASQTIIAHLIPQCIASHR 117
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 112 SWGGATLTLRLEDEAHTLSLLRRGDVLGAVTREPANVAG--CEVVELGTMRLHAIATPSL 169
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 118 RYPDIQFVLHDPQOQWVESIRQGDVDFGVIDPGVGDLCQCAI-----LSEPF 168
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 170 ----RD--AYMVDGKLDWAAMPVLRFGPKDVLQDRDLGR--VDGPVGR--RVSIVPSAE 220
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 169 LLCHRSALAVEDYVPWQALQ---GAKLVLDQYASGRPLDAALARNGTQANIV---Q 221
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 221 GFGE-----AIRGLGWLGPETQAAPMLKAGEVILLDEIPIDTPMYWQW---RLES 271
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 222 EIGHPATLFPMAAGIGISILP-ALALP-LPEGSPLVVKRI---TPVVERQLMLVLRKNR 276
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 272 SLARLTDAVVD 282
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 277 SLSTAAEALWD 287

RESULT 12
METR_SALTY
ID METR_SALTY STANDARD; PRT; 317 AA.
AC P05984;
DT 01-NOV-1988 (Rel. 09, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE TRANSCRIPTIONAL ACTIVATOR PROTEIN METR.
GN METR.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_TaxID=602;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87307964; PubMed=3040668;
RA Plamann L.S., Stauffer G.V.;
RT "Nucleotide sequence of the Salmonella typhimurium metR gene and the
RT metR-metE control region."
RL J. Bacteriol. 169:3932-3937(1987).
RN [2]
RP REVISIONS.
RA Urbanowski M.;
RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: CONTROL OF THE LAST STEP IN METHIONINE BIOSYNTHESIS;
CC METR IS A POSITIVE ACTIVATOR OF THE MET, METE AND METH GENES.
CC METR IS ALSO A NEGATIVE REGULATOR OF ITS OWN EXPRESSION.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -1- SIMILARITY: BELONGS TO THE LYSR FAMILY OF TRANSCRIPTIONAL
CC REGULATORS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M17356; AAB88753.1; -
CC StyGene; SG10227; metR.
CC InterPro; IPR000847; -
CC Pfam; PF00126; HTH_1; 1.
CC PRINTS; PR00039; HTHLYSR.
CC PROSITE; PS00044; HTH_LYSR_FAMILY; 1.
KW Transcription regulation; DNA-binding; Methionine biosynthesis;
KW Activator; Repressor.

```

FT DNA_BIND 19 38 H-T-H MOTIF (BY SIMILARITY).
SQ SEQUENCE 317 AA; 35585 MW; 90CD503F84BE8CC4 CRC64;

Query Match 9.3%; Score 136.5; DB 1; Length 317;
Best Local Similarity 23.9%; Pred. No. 0.0006;
Matches 69; Conservative 29; Mismatches 110; Indels 81; Gaps 8;

QY 6 LDPTLLSIIDEGSPEGASLISPSAVSORVKALEHHVG-RVLVSRTOPAKATEACEVLV 64
DB 7 LKTQLALNRSGSLAAAVLHQTSALSQSFDSLEORLGRFLFVRKSQLRFTPOGEVIL 66
QY 65 QAARKMVLLOAETKAQLSGRLEAIPLIITAINADSLSTWTFPPV-NEVASW-----GG 115
DB 67 OLANO-VLPQISRALQACNEPOQTRURIAIECHSCIOQLTPALENFRASWPQVEMDFTS 125
QY 116 ATLTLEDEAHTLSLLRGDVLGATVRANPVAGEV-----153
DB 126 VTFDPOPALQGGELDVMTSDILPRSELHYSMPDFEVRVLVLPADHPLASKTQITPEDLA 185
QY 154 -----VELGTMRHL--AIATPSLRDA-----YVDGKLDWAAMPVLRFGPK 192
DB 186 SEPTLLIYPQRSRLDWRHFLQAPGISPLKSVNDNTLLLIQWVAARMGTAALPHW-----240
QY 193 DVLQDRDLGRVDGPGVRRRSIVPSAEGFGEAIRRGWGLLPETQAA 241
DB 241 -----VVESVERQGLVTVTKLGDGLWSRLYAA 267

RESULT 13
LRRR_SYPN7
ID LRRR_SYPN7 STANDARD; PRT; 294 AA.
AC P52691;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE PROBABLE TRANSCRIPTIONAL REGULATOR LRRR.
GN LRRR.
OS Synechococcus sp. (strain PCC 7942) (Anacystis nidulans R2).
OC Bacteria; Cyanobacteria; Chroococcales; Synechococcus.
OX NCBI_TaxID=1140;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96281517; PubMed=8661945;
RA Anandan S., Nalty M.S., Cogdell D.E., Golden S.S.;
RT "Identification of two classes of transcriptional regulator genes in
the cyanobacterium Synechococcus sp. strain PCC 7942.";
RL Arch. Microbiol. 166:58-63(1996).
CC -!- SIMILARITY: BELONGS TO THE LYSR FAMILY OF TRANSCRIPTIONAL
REGULATORS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
CC EMBL; LA1665; AAB38541.1;
DR InterPro; IPR000847;
DR Pfam; PF00126; HTH_1;
DR PRINTS; PR00039; HTHLYSR.
DR PROSITE; PS00044; HTH_LYSR_FAMILY; 1.
KW Transcription regulation; DNA-binding.
FT DNA_BIND 18 37 H-T-H MOTIF (POTENTIAL).
SQ SEQUENCE 294 AA; 32068 MW; F9738747F149E335 CRC64;

Query Match 9.3%; Score 136; DB 1; Length 294;
Best Local Similarity 23.4%; Pred. No. 0.00061;
Matches 68; Conservative 42; Mismatches 87; Indels 94; Gaps 12;

QY 1 MNPQLDTLLSIIDEGSPEGASLISPSAVSORVKALEHHVGRVLSRTQ-PARATEA 59
DB 1 MNITQLQIFAAVETGTFSAALQLDSQSNVSRATAADELGVLLSGRGARPTRV 60
QY 60 GEVLVQAARKMVLLOAET--KAQLSGRLEAIPLIITAINADSLSTWTFPP--VFNEVASWG 114
DB 61 GERVLQAQMLQLHDSIVHEVNLKGLQGGHLRIASFRSAATHVLPPRLALFRQRCP-- 118
QY 115 GATLTLEDEAHTLSLLRRGV-LG-----AVTRE-----ANPV--- 148
DB 119 GVSVSIIETDPQGEVALREGKVDIGLLPLPRSEEDTWEITRDEVVLLPSTAHPLGQP 178
QY 149 -----ACCEVVELGTMRHLAIATPSLRDAYMYDGLDWAAMPVLRFGPK 192
DB 179 LTWAQLSRVEFTILYNVAECTT---AVRQHWATARQELKVAY-----216
QY 193 DVLQDRDLGRVDGPGVRRRSIVPSAEGFGEAIRRGWGLLPETQAA 243
DB 217 ETKEDSTI-----VSMV-----AAGLGAAILPRLAAVPI 245

RESULT 14
METR_ECOLI
ID METR_ECOLI STANDARD; PRT; 317 AA.
AC P19797;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE TRANSCRIPTIONAL ACTIVATOR PROTEIN METR.
GN METR.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=K12;
RX MEDLINE=90384950; PubMed=2205852;
RA Maxon M.E., Wigboldus J., Brot N., Weissbach H.;
RT "Structure-function studies on Escherichia coli MetR protein, a
putative prokaryotic leucine zipper protein.";
RL Proc. Natl. Acad. Sci. U.S.A. 87:7076-7079(1990).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=K12 / MG1655;
RX MEDLINE=92358234; PubMed=1379743;
RA Daniels D.L., Plunkett G.III, Burland V.D., Blattner F.R.;
RT "Analysis of the Escherichia coli genome: DNA sequence of the region
from 84.5 to 86.5 minutes.";
RL Science 257:771-778(1992).
RN [3]
RP SEQUENCE OF 1-19 FROM N.A.
RX MEDLINE=89098936; PubMed=2643109;
RA Maxon M.E., Redfield B., Cai X.-Y., Shoeman R., Fujita K., Fisher W.,
Stauffer G., Weissbach H., Brot N.;
RT "Regulation of methionine synthesis in Escherichia coli: effect of
the MetR protein on the expression of the metE and metR genes.";
RL Proc. Natl. Acad. Sci. U.S.A. 86:85-89(1989).
CC -!- FUNCTION: CONTROL OF THE LAST STEP IN METHIONINE BIOSYNTHESIS;
METR IS A POSITIVE ACTIVATOR OF THE METR, METE AND METH GENES.
CC METR IS ALSO A NEGATIVE REGULATOR OF ITS OWN EXPRESSION. BINDS
HOMOCYSTEINE AS AN INDUCER.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -!- SIMILARITY: BELONGS TO THE LYSR FAMILY OF TRANSCRIPTIONAL
REGULATORS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

```

CC  EMBL; M37630; AAA62781.1; -.
DR  EMBL; M87049; AAA67624.1; -.
DR  EMBL; AE00458; AAC76831.1; -.
DR  EMBL; J04155; AAA24159.1; -.
DR  PIR; A36066; A36066
DR  PIR; S30718; S30718.
DR  Ecogene; EG10591; metr.
DR  InterPro; IPR000847; -.
DR  Pfam; PF00126; HTH_1; 1.
DR  PRINTS; PR00039; HTHLYSR.
DR  PROSITE; PS00044; HTH_LYSR_FAMILY; 1.
KW  Transcription regulation; DNA-binding; Methionine biosynthesis;
FT  DNA_BIND 19 38 H-T-H MOTIF (BY SIMILARITY).
FT  CONFLICT 19 19 L -> T (IN REF. 3).
SQ  SEQUENCE 317 AA; 35629 MW; 513F5C710803800D CRC64;

Query Match 8.9%; Score 129.5; DB 1; Length 317;
Best Local Similarity 23.3%; Pred. No. 0.0022;
Matches 69; Conservative 33; Mismatches 111; Indels 83; Gaps 9;

QY 6 LDTLLSIIDEGSPGASLALSISPSAVSQRVKALEHHVG-RVLVSRTPAKATEAGEVLV 64
DB 7 LKTLQALRNGCSLAAATLHOTQSALSHQFSDLEQRLGFRLFVRKSOPLRFTPGGEILL 66

QY 65 QAARKMVLQAETKQALSGRLAEIPLTIAINADSLSTWFPVNEVASW-----GG 115
DB 67 QLANQ-VLPQISQALQACNEPQQTRLAIAIECHSCIQWLTPALENFHKNWPQVEMDFKSG 125

QY 116 ATLRLDEAHTLSLRGGDVLGAVTNEANPVAGCEV----- 153
DB 126 VTEDPQALQGGELDLVMTSDILPRSLGHSPMFDYEVRLVLADPHPLAAKTRITPEDLA 185

QY 154 -----VELGTRHL---AIAPSLRDA-----YMDGKLDWAAMPVLRGPK 192
DB 186 SEYLLIYPVORSRLDWRHFLQAPGVGSLSKSVNTILLIQMVAARMGIAALPHW----- 240

QY 193 DVLQDRDLGVDGPGVRRVRSIVPSAEGFEAIRRGLGMLLPETOAAFMLKAGE 248
DB 241 -----VVESPERQLVVTKTGEGLSRLYAA--VRDGE 272

RESULT 15
TFDR_ALCEU
ID TFDR_ALCEU STANDARD; PRT; 295 AA.
AC P10086; P42429;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE TRANSCRIPTIONAL REGULATORY PROTEIN TDFR/TFDS.
GN TFDR AND (TFDS OR TFDO).
OS Alcaligenes eutrophus (Ralstonia eutropha).
OC plasmid pJP4
OC Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
OC Ralstonia.
OX NCBI_TaxID=510;
RN [1]
RP SEQUENCE FROM N.A. (TFDR AND TFDS).
RC STRAIN-JMP134;
RX MEDLINE-94209236; PubMed-8157603;
RA Matrubutham U., Harker A.R.;
RT "Analysis of duplicated gene sequences associated with tfdr and tfds
in Alcaligenes eutrophus JMP134";
RL J. Bacteriol. 176:2348-2353(1994).
RN [2]
RP SEQUENCE OF 1-180 FROM N.A. (TFDS).
RC STRAIN-JMP134;
RX MEDLINE-87250253; PubMed-3036764;
RA Streiber W.R., Timmis K.N., Zenk M.H.;
RT "Analysis, cloning, and high-level expression of 2,4-
dichlorophenoxyacetate monooxygenase gene tfda of Alcaligenes

```

```

entrophus JMP134.";
RT J. Bacteriol. 169:2950-2955(1987).
RN [3]
RP IDENTIFICATION OF PROTEIN, AND POSSIBLE DNA-BINDING REGION.
RX MEDLINE-88320486; PubMed-3413113;
RA Henikoff S., Haughn G.W., Calvo J.M., Wallace J.C.;
RT "A large family of bacterial activator proteins.";
CC Proc. Natl. Acad. Sci. U.S.A. 85:6602-6606(1988).
CC -1- FUNCTION: INVOLVED IN REGULATION OF 3-CHLOROCATECHOL DEGRADATION.
CC TRANSCRIPTIONAL REGULATOR OF TDFB EXPRESSION. ACTS AS A REPRESSOR
CC IN THE ABSENCE OF ITS EFFECTOR (EITHER 2-CIS-CHLORODIENE LACTONE
CC OR CHLOROMALEYLACETATE) BUT ACTS AS AN ACTIVATOR WHEN ITS EFFECTOR
CC IS PRESENT.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -1- MISCELLANEOUS: THE PRODUCTS OF THE DUPLICATED GENE TFDR AND TFDS
CC ARE IDENTICAL.
CC -1- SIMILARITY: BELONGS TO THE LYSR FAMILY OF TRANSCRIPTIONAL
CC REGULATORS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M98445; AAA65063.1; -.
CC EMBL; M16730; -. NOT ANNOTATED_CDS.
CC InterPro; IPR000847; -.
CC Pfam; PF00126; HTH_1; 1.
CC PRINTS; PR00039; HTHLYSR.
CC PROSITE; PS00044; HTH_LYSR_FAMILY; 1.
KW Aromatic hydrocarbons catabolism; Transcription regulation;
KW DNA-binding; Activator; Repressor; Plasmid.
FT DNA_BIND 18 37 H-T-H MOTIF (POTENTIAL).
SQ SEQUENCE 295 AA; 32070 MW; C6C477BE8127A41A CRC64;

Query Match 8.8%; Score 128; DB 1; Length 295;
Best Local Similarity 24.4%; Pred. No. 0.0027;
Matches 76; Conservative 45; Mismatches 138; Indels 52; Gaps 12;

QY 5 QLTLLSIIDEGSPGASLALSISPSAVSQRVKALEHHVGRLVLSRT-OPAKATEAGEVL 63
DB 5 QLRFYFAAAEEGNVGAARLRHISQPPVTRQIHAEQHLGVLLFERSARGVQLTPAGAAF 64

QY 64 VQARKMVLQA-----ETKQALSGRLAEIPLTIAINADSLSTWFPVNE-VASWGATL 118
DB 65 LEDARRMLELGRTSVDRSRAASRGEIQ--LDIGYLGTAIYQVTPALLHAFTQAVPGATL 122

QY 119 TLRLDEAHTLSLRGGDVLGAVTNEANPVAGCEVVELGTMRLHAIATPSLRDAYMVDGK 178
DB 123 SLALMPKVRQIEALRAGTIHLGVGRFYPQEPGITVEHLHYERLYIAAGSSTAROLRQDPT 182

QY 179 LDWAAMPVLRFGPKDVLQDRDLGR---VDGPVG-RRRSVIVPSAEGFE-----AIR 227
DB 183 LLRLKSESLVLPK-----EGRPSFADEVIALMRAGVEPRVTAIVEDVNAALGLVA 234

QY 228 RGLGWGLLPETOAA---PMLKAGEVI-LLDEIPIIDTQMYQWRWLESRSRLARLTD----- 278
DB 235 AGAGTVLPASVAAIRRPFRVTMADASKVPV-----SILYLTDSRPV 280

QY 279 --AVVDAIEG 287
DB 281 LRAFLDVARRG 291

Search completed: September 13, 2001, 17:57:11
Job time: 296 sec

```


GenCore version 4.5
 Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 13, 2001, 17:56:48 ; Search time 37.95 Seconds
 (without alignments)
 1011.027 Million cell updates/sec

Title: US-09-105-117i-3
 Perfect score: 1460
 Sequence: 1 MNPIQLDRLSLIDGSEF.....RSLARLTDAVVDAIEGLRP 290

Scoring table: BLOSUM62
 Gapop 10.0 , Gapext 0.5

Searched: 425026 seqs, 132305027 residues

Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
 Maximum Match 100%
 Listing first 45 summaries

Database : SPTREMBL16:*
 1: sp_archaea:*
 2: sp_bacteria:*
 3: sp_fungi:*
 4: sp_human:*
 5: sp_invertebrate:*
 6: sp_mammal:*
 7: sp_mhc:*
 8: sp_organelle:*
 9: sp_phase:*
 10: sp_plant:*
 11: sp_rodent:*
 12: sp_unclassified:*
 13: sp_vertebrate:*
 14: sp_virus:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	590.5	40.4	300	2 Q9K4K7	Q9K4K7 streptomyce
2	472.5	32.4	303	2 O87968	O87968 streptomyce
3	410	28.1	300	2 O9HW38	O9HW38 pseudomonas
4	361.5	24.8	298	2 Q9KUN3	Q9KUN3 vibrio chol
5	177	12.1	297	2 Q9HX26	Q9HX26 pseudomonas
6	175	12.0	300	2 Q9IAJ4	Q9IAJ4 pseudomonas
7	174	11.9	295	2 Q9I085	Q9I085 pseudomonas
8	170	11.6	310	2 Q9I122	Q9I122 pseudomonas
9	163.5	11.2	286	2 Q9I435	Q9I435 pseudomonas
10	156	10.7	304	2 Q9I4D7	Q9I4D7 pseudomonas
11	152.5	10.4	292	2 Q9I0Y5	Q9I0Y5 pseudomonas
12	152.5	10.4	295	2 Q9KM14	Q9KM14 vibrio chol
13	151	10.3	302	2 Q9KNH9	Q9KNH9 vibrio chol
14	149.5	10.2	339	2 Q9HZ25	Q9HZ25 pseudomonas
15	149	10.2	303	2 Q9I1E4	Q9I1E4 pseudomonas
16	144	9.9	278	2 O86059	O86059 azospirillu
17	143.5	9.8	297	2 Q9L6S5	Q9L6S5 s 98% ident
18	142	9.7	342	2 O67999	O67999 bradyrhizob
19	141.5	9.7	311	2 Q9HXL7	Q9HXL7 pseudomonas

ALIGNMENTS

RESULT	ID	Q9K4K7	PRELIMINARY;	PRT;	300 AA.
Q9K4K7	AC	O9K4K7			
DT	01-OCT-2000	(Tremblrel. 15, Created)			
DT	01-OCT-2000	(Tremblrel. 15, Last sequence update)			
DT	01-MAR-2001	(Tremblrel. 16, Last annotation update)			
DE	PUTATIVE LYSR-FAMILY TRANSCRIPTIONAL REGULATOR.				
GN	SC5F8.17C.				
OS	Streptomyces coelicolor.				
OC	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;				
OC	Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.				
OX	NCBI_TaxID=1902;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=A3(2);				
RA	Seeger K.J., Harris D.				
RL	Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=A3(2);				
RA	Thomson N.R., Parkhill J., Barrell B.G., Rajandream M.A.;				
RL	Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.				
RN	[3]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=A3(2);				
RA	MEDLINE=97000351; PubMed=843436;				
RA	Redenbach M., Kieser H.M., Denapante D., Eichner A., Cullum J.,				
RA	Kinashi H., Hopwood D.A.;				
RT	"A set of ordered cosmids and a detailed genetic and physical map for				
RT	the 8 Mb Streptomyces coelicolor A3(2) chromosome.";				
RL	Mol. Microbiol. 21:77-96(1996).				
CC	-1- SIMILARITY: TO THE LYSR FAMILY OF TRANSCRIPTIONAL REGULATORS.				
DR	EMBL: AL357613; CAB93745.1; -				
DR	InterPro: IPR000847; -				
DR	InterPro: IPR001993; -				
DR	Pfam: PF00126; HTH_1; 1.				
DR	PRINTS; PR00039; HTHLYSR.				
DR	PROSITE; PS00044; HTH_LYSR_FAMILY; UNKNOWN_1.				
DR	PROSITE; PS00215; MITOCH_CARRIER; UNKNOWN_1.				
KW	DNA-binding; Transcription regulation.				
SQ	SEQUENCE 300 AA; 32205 MW; 7122FE5A539D642B CRC64;				

20	140	9.6	312	2	Q916X3	Q916X3 pseudomonas
21	138.5	9.5	305	2	Q9KRD6	Q9KRD6 vibrio chol
22	138.5	9.5	306	2	O86805	O86805 streptomyce
23	138	9.5	293	2	P96725	P96725 bacillus su
24	136	9.3	294	2	O9HUH6	O9HUH6 pseudomonas
25	136	9.3	298	2	O52827	O52827 rhizobium l
26	136	9.3	298	2	Q91568	Q91568 pseudomonas
27	135.5	9.3	317	2	Q916N2	Q916N2 salmonella
28	135	9.2	309	2	Q910F7	Q910F7 pseudomonas
29	134	9.2	311	2	O66393	O66393 azorhizobiu
30	134	9.2	313	2	Q9RDH0	Q9RDH0 streptomyce
31	133.5	9.1	312	2	O9L3R4	O9L3R4 anabaena sp
32	133	9.1	309	2	Q9JWC7	Q9JWC7 neisseria m
33	132	9.0	278	2	Q9RCY3	Q9RCY3 streptomyce
34	132	9.0	307	2	Q9F186	Q9F186 alcaligenes
35	131	9.0	309	2	Q9JXG8	Q9JXG8 neisseria m
36	131	9.0	314	2	O9I569	O9I569 pseudomonas
37	130	8.9	296	2	O9IAF4	O9IAF4 pseudomonas
38	130	8.9	317	2	O9HXB8	O9HXB8 pseudomonas
39	129.5	8.9	298	2	Q9I0N6	Q9I0N6 pseudomonas
40	129	8.8	292	2	Q9ZBW3	Q9ZBW3 streptomyce
41	128.5	8.8	313	2	Q9KLW6	Q9KLW6 vibrio chol
42	128	8.8	302	2	Q9I709	Q9I709 pseudomonas
43	127.5	8.7	291	2	O30441	O30441 bordetella
44	127.5	8.7	302	2	O9RH21	O9RH21 zymomonas m
45	127	8.7	306	2	Q9HY39	Q9HY39 pseudomonas

```
Query Match 40.4%; Score 590.5; DB 2; Length 300;
Best Local Similarity 44.5%; Pred. No. 1.6e-37;
Matches 129; Conservative 51; Mismatches 103; Indels 7; Gaps 5;

QY 5 QDITLIIIDEGSFEGASLALISPSAVSQRVKALHHVGRVLSRTQPAKATEAGEVLV 64
D 12 QVETLLAVDEGTFDAAAALHWTSPAVSQRVKALHRTGRVLLLTQKPVRTDGSGLV 71
QY 65 QAARKMVLLOAETKAOLSRLEIP--LTIATNADSLSTWTFPPVNEVASWGATLTLEL 122
D 72 RLARQVARLERDASAEGLRGECEPTRVTAVNADSLATWFLPALTRIPREPALCFELRR 131
QY 123 EDEAHTLSLLRGDVLGAVTREANPVAGCEVVELCTMRHLATATPSLRDAYWVDGKL--D 180
D 132 EDEGHTATLLRGVNVAAVTSPEPVGCTVTLGRMYLPDCAAPFA-ARHLDAPLRET 190
QY 181 WAAMPVLRGPKDVLQDRDLGRVDGPGVRRRVSIYPSAEGFCEAIRRGGLGGLLPETQA 240
D 191 VAGAPVVFDRDRDDFQDSFARRLGHGAGAAAR-HVYPTSEGFVEAAGLGMVGPQQA 249
QY 241 APMKAGEVI-LLDEIPIDTPMYQWRLESRLARLTDAVVDAAIEGLR 289
D 250 EPLLRTGLVTFAPDLAVDVTLYQWQMKLDSPALATVADAVVTAADALR 299

RESULT 2
ID O87968 PRELIMINARY; PRT; 303 AA.
AC O87968;
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT 01-NOV-2001 (TREMBLrel. 16, Last annotation update)
DE ORF5.
OS Streptomyces griseus.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1911;
RN [1]
RP SEQUENCE FROM N.A.
RA Uneyama T.;
RT "Open reading frame encoded around afsA gene.";
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: TO THE LYSR FAMILY OF TRANSCRIPTIONAL REGULATORS.
DR EMBL; AB011413; BAA32133.1;
DR InterPro; IPR000847;
DR InterPro; IPR001993;
DR Pfam; PF00126; HTH_1; 1.
DR PRINTS; PR00039; HTHLYSR.
DR PROSITE; PS00044; HTH_LYSR_FAMILY; UNKNOWN_1.
DR PROSITE; PS00215; MITOCH_CARRIER; UNKNOWN_1.
KW DNA-binding; Transcription regulation.
SQ SEQUENCE 303 AA; 31866 MW; FID146DD95AAOCE3 CRC64;

Query Match 32.4%; Score 472.5; DB 2; Length 303;
Best Local Similarity 39.3%; Pred. No. 1.7e-28;
Matches 125; Conservative 49; Mismatches 93; Indels 51; Gaps 12;

QY 1 MNPIDLD---TLLSIIDEGSFEGASLALISPSAVSQRVKALHHVGRVLSRTOPAKAT 57
D 1 MSELPIQVRLTLLAVDEGTFDAAAALHWTSPAVSQRVKALHRTGRVLLMTKPVRT 60
QY 58 EAGEVLQAARKMVLLOAETKAOL--SGRLAEIPLTIATNADSLSTWTFPPVNEVASWG 115
D 61 ESSEV-VALARLARLEAHEAQAALGMSGGEPTLLPIAVNSLATW-----SWQP 110
QY 116 AYLTLRLEAHTL-----SLLRRGD-VLGAVTREANPVAGCEVVELGTMR 160
D 111 CGACRRSMDSATTPGPGPYRPAAGGAGDGLVAGGDDGLFGPAA-----GMR 163
QY 161 HLAIATPSLRDAYWVDGKLDWAAM-----PVLRFQPKDVLQDRDLGRVDGPGVR--R 211
```

```
Db 164 YLPVSPGSPADRWL--GRDGTALRELIGEAPVWCFDRDDLDQAFV--RLUGPGARPSA 219
QY 212 RVSIVPSAEGFGEATRRGLGGLLPETQAAPMLKAGEVI-LLDEIPIDTPMYQWRLES 270
D 220 RRHLVPTSEGFANAVASGNGMVGMPVEQAEPLLSDGRLVRLAPEPTVDVPLVWQWKLES 279
QY 271 RSLARLTDAVVDAAIEGL 288
D 280 PALAAVAEAAVAEAAEAL 297

RESULT 3
ID Q9HW38 PRELIMINARY; PRT; 300 AA.
AC Q9HW38;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE INHIBITOR OF CHROMOSOME INITIATION ICIA.
GN ICIA OR PA4363.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PA01.
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K.; Pham X.-Q.T.; Erwin A.L.; Mizoguchi S.D.; Warrenner P.;
RA Hickey M.J.; Brinkman F.S.D.; Hufnagle W.O.; Kowalik D.J.; Lagrou M.;
RA Garber R.L.; Goltzy L.; Tolentino E.; Westbrock-Wadman S.; Yuan Y.;
RA Brody L.L.; Coulter S.N.; Folger K.R.; Kas A.; Larbig K.; Lim R.M.;
RA Smith K.A.; Spencer D.H.; Wong G.K.-S.; Wu Z.; Paulsen I.T.;
RA Reizer J.; Sailer M.H.; Hancock R.E.W.; Lory S.; Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen.";
RL Nature 406:959-964 (2000).
DR EMBL; AE004852; AAG07751.1;
DR InterPro; IPR000847;
DR Pfam; PF00126; HTH_1; 1.
DR PRINTS; PR00039; HTHLYSR.
DR PROSITE; PS00044; HTH_LYSR_FAMILY; UNKNOWN_1.
DR PROSITE; PS00044; HTH_LYSR_FAMILY; UNKNOWN_1.
SQ SEQUENCE 300 AA; 32451 MW; BFC7F298E51CFBE9 CRC64;

Query Match 28.1%; Score 410; DB 2; Length 300;
Best Local Similarity 36.9%; Pred. No. 1e-23;
Matches 107; Conservative 47; Mismatches 126; Indels 10; Gaps 5;

QY 6 LDTLLSIIDEGSFEGASLALISPSAVSQRVKALHHVGRVLSRTQPAKATEAGEVLVQ 65
D 9 LAALAAVVEQGGFERAAQAALGLSQSAVSQRIKLEARVGPVLVRETTPHPTDLGRLLN 68
QY 66 AARKMVLLOAETK---AQLSGRLAEIPLTIATNADSLSTWTFPPVNEVASWGATLTLEL 122
D 69 HVQVQVRLLEGDLQRPVNPNDDEGAPERLRIALNADSLATWAAAVGDFCAERRVLLDLY 128
QY 123 EDEAHTLSLLRGDVLGAVTREANPVAGCEVVELCTMRHLATATPSLRDAYWVDG--KLD 180
D 129 EDOEVLGKMRAGEVAGCVCGSARPVAGARSLLGLAMRYGLASPDFIARHPRGVEAA 188
QY 181 WAAMPVLRGPKDVLQDRDL-DGRVDGPGVRRRVSIYPSAEGFGEAIRRGGLGGLLPETQ 239
D 189 LAGVPAIVVEGPDLLQHRFLKDLGVEGGFIHH---LCPSSGEFVRLTAGGLGWLVPERQ 245
QY 240 AAPMLKAGEVI-LLDEIPIDTPMYQWRLESRLARLTDAVVDAAIEGL 288
D 246 VOGELARGELVELPGQVIDVPLVYWHYWRNGELLASLTHELLARAGDGL 295

RESULT 4
Q9KUN3
```

RN	SEQUENCE FROM N.A.					
RP	STRAIN=PA01;					
RC	MEDLINE=20437337; PubMed=10984043;					
RX						
RA	Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,					
RA	Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,					
RA	Garber R.L., Goltzy L., Tolentino E., Westbrook-Wadman S., Yuan Y.,					
RA	Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,					
RA	Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,					
RA	Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.;					
RA	"Complete genome sequence of Pseudomonas aeruginosa PA01, an					
RT	opportunistic pathogen."					
RL	Nature 406:959-964(2000).					
DR	EMBL; AE004817; AACG7382.1; -.					
DR	InterPro; IPRO00847; -.					
DR	Pfam; PF00126; HTH_1; 1.					
DR	PROSITE; PS00044; HTH_LYSR_FAMILY; UNKNOWN.1.					
SQ	SEQUENCE 297 AA; 32515 MW; 3809B555DD453B720 CRC64;					
	Query Match	12.1%;	Score 177;	DB 2;	Length 297;	
	Best Local Similarity	28.0%;	Pred. No. 6.7e-06;			
	Matches 78;	Conservative 38;	Mismatches 119;	Indels 44;	Gaps	
QY	4	IQLDTLLSIIDEGFEGASLALSISPSAVSQRVKALEHHVGRVLVSR-TOPAKATEAGEV 62				
Db	14	VLLTRFELEVDSSGGFAALADNLATLTPSAVSGHIKRLQTAGVSLLSRTTTRRLLELTQAGET 73	:	:	:	:
QY	63	LVOAARKWVLQAEETHKAOLSGRLAEIPLTTAINADSLSLTFWPPPVFNVEVASW----- 113	:	:	:	:
Db	74	LYAYARNIVELEREARLRHGHTPIRDRLRGASEDFASAWLPRLVQRFRKHHPPEASIELK 133	:	:	:	:
QY	114	-GGATTLRLREDEAHT-----LSLLRRGDVLGAVTREANPVACEVVVELCTMRHLAIA-- 165	:	:	:	:
Db	134	VGITTDLLRQAQAGRTDVFVKQCRRVRDGDGELLWEELPWAAATAIEPAGEPLPALF 193	:	:	:	:
QY	166	-TPSL-RDAYMVDGKLDAAMPVFLRFPGKDVLOQDRDLGDRVDPGVGRRRSYIVSPSABGF 223	:	:	:	:
Db	194	PEPCVYREAII-ALGAARPW-----RLVFESAMAGCLSAALAGFAVTTVVARSQ--- 243	:	:	:	:
QY	224	EATRGL-----GWGLLPETQ-----AAPMLKAGEVI 250	:	:	:	:
Db	244	--MREGRLRELCPQGFPPTLPPEARFYAFSPQPSLAADALI 280	:	:	:	:
RESULT	6					
Q914J4	ID	Q914J4	PRELIMINARY;	PRT;	300 AA.	
AC	Q914J4;					
DT	01-MAR-2001	(TEMBLrel. 16, Created)				
DT	01-MAR-2001	(TEMBLrel. 16, Last sequence update)				
DT	01-MAR-2001	(TEMBLrel. 16, Last annotation update)				
DE	PROBABILE TRANSCRIPTIONAL REGULATOR.					
GN	PA1141.					
OS	Pseudomonas aeruginosa.					
OC	Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;					
OC	Pseudomonas.					
OX	NCBI_taxid=287;					
RN	[1]					
RP	STRAIN=PA01.					
RC	MEDLINE=20437337; PubMed=10984043;					
RX						
RA	Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,					
RA	Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,					
RA	Garber R.L., Goltzy L., Tolentino E., Westbrook-Wadman S., Yuan Y.,					
RA	Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,					
RA	Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,					
RA	Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.;					
RA	"complete genome sequence of Pseudomonas aeruginosa PA01, an					
RT	opportunistic pathogen."					
RL	Nature 406:959-964(2000).					
DR	EMBL; AE004544; AAG04530.1; -.					
DR	InterPro; IPRO00847; -.					

Qy	6	LDLTLSIIDSGSPGASLALSISPSAVSQRVKALEHHVGRVLVSRTPQAKATEAGEVLV	64
Db	8	IELFLAVLDISGSGAARALGRVPSVSMGIANLEAEALFALFERTHREARPTALARALA	67
Qy	65	QAARKWVLLQAETKAOQLSGRLAETP-----LITAINAD-----SLSTWPPPVF	107
Db	68	POARMI-----AEQLAQVQVIALELSQGLSRSLGVGVADDSGRFGALFTLSERYPLLD	123
Qy	108	NEVASWGATLTURLDEAHTLSLLRRGDV---LGAVNREANP-----VAGCEVVELGT	158
Db	124	VEVLS-----AAQDDA--LALLHGGRISLFCFGAGISVNPQERFQYVGAESLVATIS	173
Qy	159	MRHLAIATPSLRDAYMYVDKGLDWAAMPVLRFGPKDVLQDRD--LDGRVDGPGVRRRYSIVP	217
Db	174	PRHPALQKPG-QALYL-----EELVNRQIILVASCDDLPLADTRPLIAG	215
Qy	218	S---AEGFG---EAIIRGLGWGLLPETQAAPMLKAGEVILLD-----EIPIDTPMYQR	265
Db	216	ACWRTDSLGTALFMEVGAIGWGNFPLSRVAPLLATGRVLRLDFRNTKNELKLPVHAIW--	273
Qy	266	WRLESRLARLTDVAV	281
Db	274	--LKNQPLRKAAGELV	287
RESULT	8		
Q91122	AD	Q91122	PRELIMINARY; PRT: 310 AA.
IC	Q91122		
DT	01-MAR-2001	(Tremblrel. 16, Created)	
DT	01-MAR-2001	(Tremblrel. 16, Last sequence update)	
DT	01-MAR-2001	(Tremblrel. 16, Last annotation update)	
DE		PROBABLE TRANSCRIPTIONAL REGULATOR.	
GN	PA2123		
OS	Pseudomonas aeruginosa.		
OC	Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;		
OC	Pseudomonas.		
OX	NCBI_TaxID=287;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=PA01.		
RA	MEDLINE=20437337; PubMed=10984043;		
RA	Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warriner P.,		
RA	Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,		
RA	Garber R.L., Goltzy L., Tolentino E., Westbrook-Wadman S., Yuan Y.,		
RA	Brody L.L., Coulter S.H., Folger K.R., Kas A., Larbig K., Lim R.M.,		
RA	Smith K.A., Spencer D.N., Wong G.K.-S., Wu Z., Paulsen I.T.,		
RA	Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.;		
RT	"Complete genome sequence of Pseudomonas aeruginosa PA01, an		
RT	opportunistic pathogen."		
RL	Nature 406:959-964 (2000).		
DR	EMBL; AE004639; AAG05511.1;		
DR	InterPro; IPR000847;		
DR	Pfam; PF00126; HTH.1;		
DR	PRINTS; PR00039; HTHLYSR.		
SQ	SEQUENCE 310 AA; 34611 MW; 258EL5C239BEFE7C CRC64;		
Query Match	11.6%;	Score 170;	DB 2; Length 310;
Best Local Similarity	25.7%;	Pred. No. 2.4e-05;	
Matches	77; Conservative	37; Mismatches 98;	Indels 88; Gaps
Qy	5	QLDTLLSIIDGSPGASLALSISPSAVSQRVKALEHHVGRVLVSRTPQAKATEAGEVLV	64
Db	6	EMTTFVSVDAHSPFSAARLGTGTSQVSQRIQQLERRLGRLLNRTPLSLTDPGRTY	65
Qy	65	QAARKWV--LQAET-----KAQLSGRLAETP-----LITAIN	95
Db	66	EHACRLQLEEQEAASVLDADDLRLGRLSAPLAFTPYRLAPLLARFAERHQLRVQDVQ	123
Qy	96	ADS--LSTWFFPVFNEVASWGATLTURLDEAHTLSLLRRGDVLGAVTREANPVGCEV	154
Db	126	ADDFVNLOSPRFDMAIRMG-----RLDDSLV-----ARPT-----	158

DT	01-MAR-2001 (TReMBLrel. 16, Created)	
DT	01-MAR-2001 (TReMBLrel. 16, Last sequence update)	
DT	01-MAR-2001 (TReMBLrel. 16, Last annotation update)	
DE	PROBABLE TRANSCRIPTIONAL REGULATOR.	
GN	PA1201.	
OS	Pseudomonas aeruginosa	
OC	Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;	
OC	Pseudomonas.	
OX	NCBI_TaxID=287;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=PA01.	
RA	MEDLINE=20437337; PubMed=10984043;	
RA	Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrenner P.,	
RA	Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,	
RA	Garber R.L., Goltzy L., Tolentino E., Westbrook-Wadman S., Yuan Y.,	
RA	Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,	
RA	Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,	
RA	Reizer J., Salier M.H., Hancock R.E.W., Lory S., Olson M.V., an	
RT	"Complete genome sequence of Pseudomonas aeruginosa PA01, an	
RT	opportunistic pathogen."	
RL	Nature 406:959-964(2000).	
DR	EMBL: AE004550; AAG04590.1; -.	
DR	InterPro: IPR000847; -	
DR	Pfam: PF00126; HTH_1; 1.	
DR	PRINTS: PR00039; HTHLYSR.	
DR	PROSITE: PS00044; HTH_LYSR_FAMILY; UNKNOWN.1.	
SQ	SEQUENCE 304 AA; 33974 MW; 9AA2E242AE5CA8B1 CRC64;	

Query Match	10.7%;	Score 156;	DB 2;	Length 304;
Best Local Similarity	26.7%;	Pred. No. 0.00028;		
Matches	76;	Conservative	52;	Mismatches 109; Indels 48; Gaps

Qy	6	LDTLLSIDESPGASLALSISPSAYSORVKALEHHVG-RVLVSRTPQAPKATEAGEVLV 64
Db	8	IOAFNLNVVLEISAAERMSLSKSVISKRVSLDERHLGVRLLYRSTRNVEPTEAGGFFY 67
Qy	65	QAARKWVLQ-----AETKA---QLSGRLAEI-PLTAINADSLSTWFFPPVFNE-VASW 113
Db	68	KSAR--ASLQDLNNAESVALRENDLCELVRVAPMSFG-----TLWLGLYMEFMARN 119
Qy	114	GGATLTURLEDEHTLSLLRRG-DVLGAVTR--EANPVAGEVVELGTMRHLAIPSLR 170
Db	120	PRLEVLQLDD--RIVDFEKEGYDLAIRITRLQDSSLIAR---QLGTSRRVCCSPEYL 173
Qy	171	DAYWVDGKL---DWAAMPVLRFCPKVDLDRLDGRVDG-----PVGRRRVSVIPA 219
Db	174	ERH---GFLQRIEDILCHPCIGYSHNTPSQLWSEFPVAGEPARMTPRGRFNTN---NG 277
Qy	220	EGFGEATRRGLGWLLPETQAAPMLKAGEVI--LLDEIPIDTPMY 262
Db	228	QTMREDAVRGLGLAWLPLFIAAEDLAAGRLVEALPQERPLDLLIY 272

RESULT	11
Q910Y5	ID
Q910Y5	PRELIMINARY; PRT; 292 AA.
AC	Q910Y5;
DT	01-MAR-2001 (TReMBLrel. 16, Created)
DT	01-MAR-2001 (TReMBLrel. 16, Last sequence update)
DT	01-MAR-2001 (TReMBLrel. 16, Last annotation update)
DE	PROBABLE TRANSCRIPTIONAL REGULATOR.
GN	PA2497.
OS	Pseudomonas aeruginosa.
OC	Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC	Pseudomonas.
OX	NCBI_TaxID=287;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=PA01.
RA	MEDLINE=20437337; PubMed=10984043;
RA	Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrenner P.,

RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Golltry L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Sater M.H., Hancock R.E.W., Lory S., Olson M.V.,
RT *Complete genome sequence of *Pseudomonas aeruginosa* PA01, an
RT opportunistic pathogen.
RL Nature 406:959-964(2000).
DR EMBL: AE004677; AAG05885.1; -.
DR InterPro: IPR000847; -.
DR Pfam: PF001066; -.
DR PRINTS: PR00039; HTH_1; 1.
DR PROSITE: PS00044; HTH_LYSR_FAMILY; UNKNOWN_1.
DR PROSITE: PS00216; SUGAR_TRANSPORT_1; UNKNOWN_1.
SQ SEQUENCE 292 AA; 32054 MW; 1B7E8AB9C4396CDF CRC64;

Query Match 10.4%; Score 152.5; DB 2; Length 292;
Best Local Similarity 25.9%; Pred. No. 0.00049;
Matches 84; Conservative 41; Mismatches 122; Indels 77; Gaps 14;

QY 4 IQDITLLSIIDSGSPGASLALISPSAVSORVKALEHHVGRVLSRTQ-PAKATEAGEV 62
DB 5 VILRTFVSVDVDTGNFRAGEHLHLTQSTVSQQVIRLEQNLGCRLLDRSQRVLPTEGER 64
QY 63 LVQAARKVLLQAETKAKQLSGLAEIPLTAINADSLSTWFFPVEVNEVASWGATLRL 122
DB 65 LLGYARRLLRLEDEASEALSPAHGDLVGLVPEDLAGEVLPVLTFRTE---ERPLRL 121
QY 123 EDEA---HTLSLLRGDVLGAVTREANPVAGCEVVELGTMRHLAATPSLRDAYMVDGK 178
DB 122 EVESGLSHLLRLYRSGEDLLLVQWQADSDC-----HARWAE----- 161
QY 179 LDW---AAMPVLRFGPKVQLDRDLGRVDGPGVR-RRVSIIVPSAEGFCE----- 224
DB 162 LGWFGSAARPFGESEEPVP-----LVVFGVLYRQEMITHALESIGRRWRIGYSSAS 215
QY 225 -----AIRRGLGWLLP-----ETQAA-PMLKAGEVILLDEIPDTPMWRW 266
DB 216 LASLVAAVAGVGLVSLPLGCVGPEHRLGAGAGPPPIAGLELALYARPELDS----- 268
QY 267 RLESRLA-RLTDVAVDAIEGLR 289
DB 269 --AGRTLRLDRD-LGDARLEGILQ 289

RESULT 12
Q9KM14 PRELIMINARY; PRT; 295 AA.
AC Q9KM14;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)
DE TRANSCRIPTIONAL REGULATOR, LYSR FAMILY.
GN VCA0575.
OS *Vibrio cholerae*.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; *Vibrio*.
OX NCBI_TaxID=666;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=EL TOR N16961 / SEROTYPE O1;
RX MEDLINE=20406833; PubMed=10952301;
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
RA McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA Fraser C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen *Vibrio*
RT *cholerae*.";
RL Nature 406:477-483(2000).

DR EMBL: AE004388; AAF96477.1; -.
DR TIGR: VCA0575; -.
DR InterPro: IPR000847; -.
DR Pfam: PF00126; HTH_1; 1.
DR PROSITE: PS00044; HTH_LYSR_FAMILY; UNKNOWN_1.
SQ SEQUENCE 295 AA; 32847 MW; 4BBD85F2A87B2F51 CRC64;

Query Match 10.4%; Score 152.5; DB 2; Length 295;
Best Local Similarity 25.5%; Pred. No. 0.0005;
Matches 79; Conservative 62; Mismatches 102; Indels 67; Gaps 17;

QY 1 MNPIOLDTLLSIIDSGSPGASLALISPSAVSORVKALEHHVGRVLSRTQ-PAK 55
DB 2 LSPITLEA-LHILDAIERRGSFAAANELNRPSSLSYQIQLEQDLDMIFDRSGHRAN 60
QY 56 ATEAGEVLVQAARKVLLQAETK-----AQLSRLAEIPLTAINADSLSTWFFP---VFN 108
DB 61 FTEAGKLLIERGR--AILAATEKLVNDATLLANGWELDTITIALDG-----IVPAANLFP 112
QY 109 EVASWGATLT-LRLEDE--AHTLSLLR--RGDVLGAVTREANPVAGCEVVELGTMRHLA 163
DB 113 MVEALGNISKTFRVQDEILACGWEALATGRADLLICPRIEALP-QDYKAETIGTMKMIW 171
QY 164 TATPS-----LRDAY-----MVDGKLDWAAMPVLRFGPKVQLDRDLGRVDG 206
DB 172 VAAPTHYVRRSGEFNEAREKYRAIATADTAREQPAMSV-----NILQ----- 215
QY 207 PVRRRSIVPSAEGFGEAIRGLGWLLPETQAAPMLKAGEVILL---DEIPDTPMWRW 263
DB 216 ---ROPRTVSNLDAKCALVAGLGTGLPLQVAPYIDKGLKAIHGSDELMEDIVLAW 272
QY 264 QRWRL-ESRS 272
DB 273 RNNQGEAKS 282

RESULT 13
Q9KNH9 PRELIMINARY; PRT; 302 AA.
AC Q9KNH9;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)
DE TRANSCRIPTIONAL REGULATOR, LYSR FAMILY.
GN VC2760.
OS *Vibrio cholerae*.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; *Vibrio*.
OX NCBI_TaxID=666;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=EL TOR N16961 / SEROTYPE O1;
RX MEDLINE=20406833; PubMed=10952301;
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
RA McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA Fraser C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen *Vibrio*
RT *cholerae*.";
RL Nature 406:477-483(2000).
CC -!- SIMILARITY: TO THE LYSR FAMILY OF TRANSCRIPTIONAL REGULATORS.
DR EMBL: AE004342; AAF95899.1; -.
DR TIGR: VC2760; -.
DR InterPro: IPR000847; -.
DR Pfam: PF00126; HTH_1; 1.
DR PRINTS: PR00039; HTH_LYSR.
KW DNA-binding; Transcription regulation.
SQ SEQUENCE 302 AA; 34281 MW; DDEEB28511225CE9 CRC64;

QY	7	DTL-----LSIIDEGSPGASLALISPSAVSQRVKALEHHVGRVLSVRTQPA-KATE	58
		: : : : : : : : : : : : :	
Db	25	DTLSEQFVLVDLDVGDSFAAAKKHPLTPSAVARMDALERAVGSTLLVTRTTHAVRATP	84
QY	59	AGEVILVOARKWV----LIQAEYKAQLSG--RLAETPLTIAINADSLSTWFFPPVFNVEVAS	112
		: : : : : : : : : : : : :	
Db	85	AGLAFADARRIVTELRLARAEAVSLSTAPQGLIRIDAPVPGRRHLA---PAVADFLKA	141
QY	113	WGGATITLRLEDAHYLSLLRRCGVIGAV--TREANPVAGCEVW--ELGTMRHRLATATPS	168
		: : : : : : : : : : : : :	
Db	142	NPGLDVLQRLIGSFIDL---QGEHLGEVDVLVRACGLPDSRLVATSLAPMVRVVCASPE	197

AC Q9IE4:
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE PROBABLE TRANSCRIPTIONAL REGULATOR.
GN PA2334.
GN

QY	5	QLDTLSIIDEGSFEGASLALSIPSAVSORVKALEHHVGRVILSVR-TQPAKATEAGEVL	63
Db	7	LRHFETALHEHRSFVAAALAVNLSSQAFSRISQIALEHSAGCRVLDRAAKDLRPTKQGLIV	66
QY	64	VQARKMVLVLAQETKAQLSG--RLA-EIPLTIAINADSL-----STWFPP-	105
Db	67	LEHARLV-----SCAHNLANEIALFNGELREFGAGPATWLVLPVAVGREN	116
QY	106	-----VFNEVASWGATVLTLRLE-----DEAHTLSLL--RGCDVLGAVT	142
Db	117	VRYPKARVSQEVHQAQALNRRLQAAEEFFVADTRHFEADPOVYTRSLRPRRWHF--CC	173
QY	143	REANPVAGCWEVLGTMRHLAIAI-----PSLRDAYWVDGKLDWAAMPVLRFGPKVDLQDR	198
Db	174	REGHPLAARESVAEELLSPLATNISPPNIRKVLV-----	209
QY	199	DLGRVD-GPV-----GRRRVSVPSAEFGGEAIRGLGWGLLPTQAAAPMLKAG-EVIL	251
Db	210	DLSGRDFPRAIECENGHSLUGVLLASDAITG--ISSGLG-----DLAPPGGGGLRUPR	260
QY	252	LDEIPTDTPMYQWRWRLESR-----SLARLTDAVVD	282

Db 261 VRELDDDELHTRYGIVSRAGYSLPLAAMID 294

Search completed: September 13, 2001, 17:56:49
Job time: 299 sec